Machine Learning and Forward Looking Information in Option Prices

Hu, Qi

Award date:
2018

Awarding institution:
University of Bath

Machine Learning and Forward Looking Information in Option Prices

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

submitted by

Qi Hu

September, 2018

[Signature]______________________________

With great respect and appreciation,

I would like to dedicate this thesis to my family, who encourage and support me all the time.

I also would like to dedicate it to Professor David P. Newton who made this work possible.

Acknowledgements

Undertaking this PhD has been a long and life-changing experience for me. I could not have completed this without the support and encouragement from many people who I deeply appreciate in various ways.

First and foremost, I would like to thank my supervisor Professor David P. Newton. Without his encouragement and support, I probably would have given up along my PhD journey. His continuous emotional support, invaluable suggestions and friendship help me through the tough times. Also, I want to thank Professor Michael Tretyakov and Professor Ania Zalewska for their excellent guidance and supervision.

Further, I want to express my respect to Professor Stephen Ross, who proposed The Recovery Theorem. He will be remembered and missed by me. The implementation of The Recovery Theorem (or Ross Recovery) was a good start in my research. It has helped me to understand option pricing theory and asset pricing model fundamentally.

In addition, I am indebted to many of my colleagues for their friendship and support. I would like to thank: Soap (Haozhe Su), Tony (Hui Tian) and Harry (Yulin Wu). They are wonderful friends and the name order is random. Moreover, special thanks go to staff in the University of Bath and University of Nottingham. I am grateful for their help. I thank the University of Bath, particularly Professor Bernie Morley, for a tuition fee scholarship.

Last but not the least, I would like to thank my family, who always support and love me. My PhD journey would be difficult without your love and support. I would like to express my deep appreciation to my husband Liang Liu, for his patience and continuous support.

Table of Contents

Appendices

List of Tables

List of Figures

Abstract

The use of forward-looking information from option prices attracted a lot of attention after the 2008 financial crisis, which highlighting the difficulty of using historical data to predict extreme events. Although a considerable number of papers investigate extraction of forward-information from cross-sectional option prices, Figlewski (2008) argues that it is still an open question and none of the techniques is clearly superior.

This thesis focuses on getting information from option prices and investigates two broad topics: applying machine learning in extracting state price density and recovering natural probability from option prices. The estimation of state price density (often described as risk-neutral density in the option pricing literature) is of considerable importance since it contains valuable information about investors' expectations and risk preferences. However, this is a non-trivial task due to data limitation and complex arbitrage-free constraints. In this thesis, I develop a more efficient linear programming support vector machine (L_1-SVM) estimator for state price density which incorporates no-arbitrage restrictions and bid-ask spread. This method does not depend on a particular approximation function and framework and is, therefore, universally applicable. In a parallel empirical study, I apply the method to options on the S&P 500, showing it to be comparatively accurate and smooth. In addition, since the existing literature has no consensus about what information is recovered from The Recovery Theorem, I empirically examine this recovery problem in a continuous diffusion setting. Using the market data of S&P 500 index option and synthetic data generated by Ornstein–Uhlenbeck (OU) process, I show that the recovered probability is not the real-world probability. Finally, to further explain why The Recovery Theorem fails and show the existence of associated martingale component, I demonstrate a example bivariate recovery.

Notation and Abbreviations

Items	Description
$(\Omega, \mathcal{F}, P)$	Probability space
S	Underlying asset price, S_t is the price at time t, S_T is the price at maturity date
r	Interest rate, in option pricing theory, it can be assumed as constant, deterministic function or zero
T	Option maturity date
K	Option strike price
F^{τ}	Forward Price
$Z(S_T)$	Option payoff
k	Forward moneyness $\frac{k}{F^{\tau}}$
τ	Time to maturity of Option, it usually is calculated as $\frac{T-t}{365}$
σ	Volatility of underlying
$C(k, \tau)$	Call option price with strike k and time to maturity τ
$P(k, \tau)$	Put option price with strike k and time to maturity τ
ITM	In the money, which refers strike price is less than underlying price
ATM	At the money, which refers strike price is equal than underlying price

Continued on next page

Items	Description
OTM	Out the money, which refers strike price is great than underlying price
SDF	Stochastic discount factor
EMH	Efficient market hypothesis
ARCH	Autoregressive conditional heteroskedasticity
GARCH	Generalized autoregressive conditional heteroskedasticity
AI	Artificial intelligence
AIC	Akaike Information Criterion
SVM	Support vector machine
QP	Quadratic Programming
LP	Linear Programming
$\mathbb{P}$	Real-world probability measure, probability density under $\mathbb{P}$ is $f(S_T)$
$\mathbb{Q}$	Risk neutral measure, the risk neutral density is $p^*(S_T)$, risk neutral distribution is $F(S_T)$
$\mathbb{Q}^{\mathbb{T}}$	Forward measure
$E_t(.)$	Expectation under real-world probability
$K(x,x)$	Kernel function
$\mathbf{k}$	Vector of forward moneyness
$\boldsymbol{\tau}$	Vector of time to maturity
α_i	Coefficient of L_1-SVM approximation
C,λ	Trade-off parameters balance the estimated error and no arbitrage conditions
a,ε,ξ	Bound parameters of L_1-SVM approximation

Continued on next page

Items	**Description**
$r_k(x)$	Derivatives vector of kernel function
$Y_k(Z_k)$	Vector of target value y
$\otimes$	Kernel product of two matrix
S	State Price matrix
P	Real-world transition matrix
Q	Transition matrix

Chapter 1

Introduction

In this chapter, I outline the research background, motivations, and main contributions and findings. Section 1.1 briefly introduces the history of options. Section 1.2 summaries research gaps and motivation. An overview of the main findings of this thesis is presented in Section 1.3. Section 1.4 provides the thesis structure.

1.1 Overview and Background

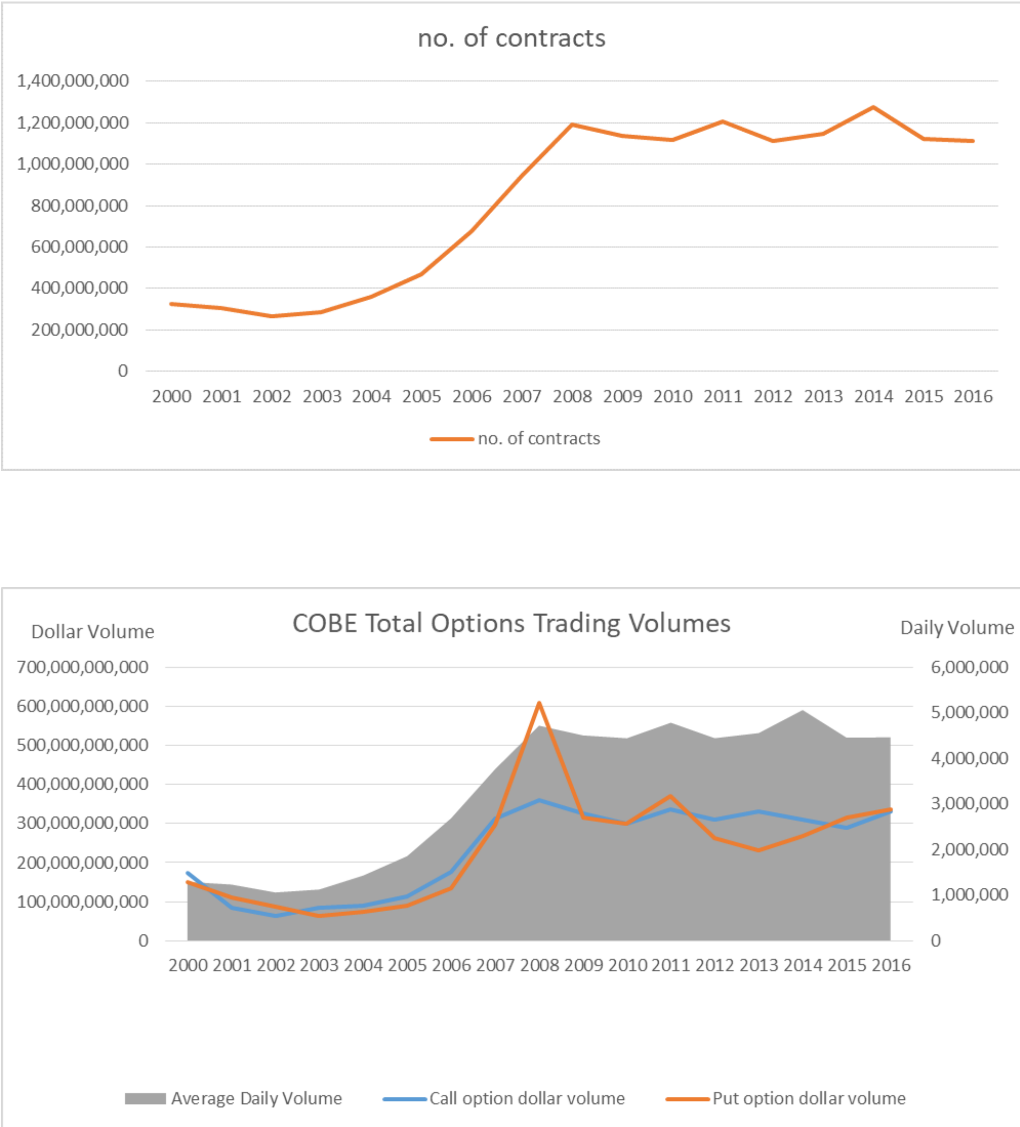
> We live in a world full of contradiction and paradox, a fact of which perhaps the most fundamental illustration is this: that the existence of a problem of knowledge depends on the future being different than the past, while the possibility of the solution of the problem depends on the future being like the past.
>
> —Knight (1921)[1]

An option is a financial contract that gives a buyer (or a seller) a right to buy (or sell) a certain amount of something which I name the underlying at specific price on a specified date[2]. The earliest example of using options can be traced

[1] Page 313

[2] The key insight of option contact is: it gives the right but not the an obligation to exercise contract. This intuition is easy to understand as instead of an obligation, people always have a choice (option) in real life

Figure 1.1.1: Number of Contracts and Trading Volume on CBOE



Note: The top panel plots the number of contracts traded on CBOE from 2000 to 2016. The bottom panel reports the average daily trading volumes and dollar volume of call and put options on CBOE in the same period. The orange line represents the dollar volume of call option while the blue line plots the dollar volume of put options. The gray areas plot shows the average daily volume changes. The data is available at http://www.cboe.com/data/historical-options-data/annual-market-statistics back to ancient Greece[3], when Thales of Miletus made a successful speculation by designing an olive presses option contact. According to Aristotle [4], Thales could predict weather conditions. Using this information, Thales successfully forecasted a rich harvest of olives in a particular year and paid a low deposit (option premium) to pre-book all olive presses in advance. This deposit gave him the right to use the olive-presses if he so choose. When the season came and the demand of olive-presses surged, he rented the olive presses to farmers at a higher price (excising his options). This is a classic example of using call options to make a profit. However, not all people seeing similar opportunities got lucky. Thompson (2007) remarks that tulip-mania in 1637 was associated with the use of options. On 24th February 1637, Dutch florists announced a new trading rule which transformed the tulip-bulb forward contract into an option contract. Under the new rule, investors did not need to pay in the forward or spot market to buy tulip bulbs. They could pay only 3 percent of the forward contract price to reserve a right (without obligation) to buy tulip bulbs in the future. If the market soared, they profited like Thales of Miletus. If the market slipped, they would have lost just 3 percent. This announcement resulted in a massive rise in tulip options prices and many people invested their lifes' savings in tulip options. In the next few months, the strike price of tulip option traded as high as 10 times the spot price. Then the bubble burst and many went bankrupt as the tulip fever suddenly ended. This tulip disaster set a bad image of options over decades[5]. For a century, options trading was illegal in Great Britain. On the other hand, options trading in the US began and develop rapidly afterward. In 1872, an American

[3]Some researchers believe that the story (the marriage between Jacob and Laban's daughter) in The Bible is the first record option transaction

[4]Please refer to Aristotle (1999) for more detail. The original record is in Aristotle, Politics 1.1259a, Book I, Chapter 11, sections 5-10

[5]The other famous derivative disaster is Baring Bank in 1995, Nick Lesson traded aggressively in Nikkei 225 and SIMEX futures and options and lost 827 million pounds.

businessman, Russel Sage, created call and put options. In 1973, a remarkable year in option history, the first official traded options market - Chicago Board Options Exchange (CBOE) was opened and Fischer Black, Myron Scholes and Robert Merton(Black and Scholes (1973); Merton (1973)) published their option valuation model. Since then, the option market has grown explosively. Exchanges were opened all over the world and products such index option, spread option, VIX options and Long-term Equity Anticipation Securities (LEAPS) began to be traded.

Nowadays, with introduction of electronic trading platforms and increased computation power, the trading of options has become more popular and easier. According to the Bank for International Settlements(BIS)[6] and the Futures Industry Association(FIA)[7], traded option contracts account for 10% of the global OTC (over-the-counter; i.e, not traded on an exchange but directly) derivative market and 40% of exchange traded contracts in 2017. Furthermore, both the number of tradable option contracts and trading volumes have soared since 2005. As shown in Figure 1.1.1, the number of options contracts has climbed from less than 400 million to more than 1200 million from 2000 to 2016. In particular, the number of contracts increased dramatically from 2005 to 2006. This significant change corresponds with the introduction of weekly, short-term options and VIX options in 2005 and 2006[8]. In addition, the daily trading volume and dollar volume maintained an historically high level during the past 16 years with more than 4 million options traded daily on the CBOE over 8 years.

[6]see https://www.bis.org/statistics/derstats.htm for more detail
[7]see https://fia.org/categories/exchange-volume for more detail
[8]see http://www.cboe.com/aboutcboe/history for more detail

1.2 Motivations and Research Gaps

This thesis focuses on extracting information from options market data: more precisely, the risk neutral density(RND), underlying natural probability and investors' risk preferences. As the earlier section has described, both Thales of Miletus and Dutch saw a similar money-making opportunity. However, one profited but the others' hopes ended badly. The main difference between them is underlying movement and investor's risk preference. Obviously, if any investors have true foresight about either item, they could make handsome profits (or avoid unnecessary speculation). However, this is impossible in real markets. How to forecast future underlying movement has been a fundamental concern of both scholars and practitioners for many years. There are numerous business, equity, financial analysts in Wall Street writing market outlook reports and thousands of researchers using sophisticated financial forecasting models every day. As Granger (2005) states "more likely the results are fragile, once you try to use them, they go away".

From time series econometrics and rational expectations perspective, scholars and practitioners always try to draw on insights from historical data (such as using autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) model). However, this approach is directly challenged by prominent economists like Knight (1921)[9] and the famous efficient market hypothesis (EMH) by Fama (1970). Fama (1970) argues that all available information in the market has already been reflected in the current asset price. Future asset price movements are only influenced by new information. Knight's and Fama's comments gained renewed interest after 2008 when conventional risk management techniques have been questioned for

[9] see quote before this chapter

their inability to manage tail risk and black swan events. Historical time series data do not offer many guidelines in predicting extreme events. Recently researchers have shifted their attention to inferring underlying probability densities from options markets. As an option is a forward-looking instrument, the benefits of focussing on cross-sectional option data seem obvious in principle. First, an option contains information on expectations about underlying future movement. Second, unlike traditional time series econometrics, inferring probabilities from option prices requires no prior guesses about the underlying distribution.

Starting with the remarkable Black-Scholes-Merton model (Black and Scholes (1973); Merton (1973)), a large body of option pricing literature has developed to calculate the fair value of an option price. Wilmott (2006) classifies the option pricing methods into four catalogues: Lattice(tree) method (Cox et al. (1979)), Monte Carlo method (Boyle (1977)), Finite Difference method (Brennan and Schwartz (1977)) and Numerical Integration (Andricopoulos et al. (2003, 2007); Chen et al. (2014))[10]. All of these are based on the risk neutral measure $\mathbb{Q}$. Following this theoretical framework, a number of studies have concentrated on developing techniques to extract the risk neutral probability density (RND) from option prices (see Figlewski (2018) and Chapter 5 for a brief review), but none shows superior performance. This is not a surprise since estimating the risk neutral density poses five challenges (see Chapter 5 for more detail).

- The estimation of RND is based on continuous strike price while the strike price in the real market is discrete
- The estimation of RND ignores the information contained in the bid-ask spread

[10]Because this thesis focus on the inverse problem of option pricing(a.k.a estimate underlying probability density from option price), I recommend reader refer Chapter2 and textbooks Wilmott (2006) and Hull and Basu (2016) for more information.

- Market option data contains noise from various sources which may lead to multimodal RND
- Theoretical RND lies in $[0, \infty]$ while the market option data can only estimate RND within bounds
- The estimation of RND suffers the 'curse of differentiation'.

Unlike the estimation of RND, how to separately recovery underlying real-world probabilities from option prices is a new topic. Prior to Ross (2015)[11] , who proposes The Recovery Theorem (hereafter TRT), there is no model-free method that could uniquely recovery real-world probabilities of underlying from option prices[12]. This break-through work seems to challenge the traditional option pricing framework under the risk neutral measure $\mathbb{Q}$ and provide a new insight to re-examine or build all pricing, risk management frameworks under the recovered probability. However, until now, existing studies have reached no consensus on what information is recovered by Ross's theory. The original TRT provided by Ross has argued that the recovered probability is the real-world probability, namely the $\mathbb{P}$ probability of the underlying asset. This argument is theoretically supported by Carr and Yu (2012), who provide an alternative way to derive TRT from a numeraire portfolio. However, Borovička et al. (2016) point that the since there is always a martingale component contained in the recovered probability, the recovered probability is between $\mathbb{Q}$ and $\mathbb{P}$ (They call this probability as natural probability).

Although researchers and practitioners have demonstrated a strong interest in using forward-looking information from option prices, two main problems are

[11] The working paper is available online in 2012.

[12] To my best knowledge, there are only several studies try to solve this problem assuming particular stochastic discount factor(Jackwerth and Rubinstein (1996))

associated with extracting this information.

- How to extract a well-behaved RND from option prices?
- What information do I recovery from TRT, the real-world probability or natural probability?

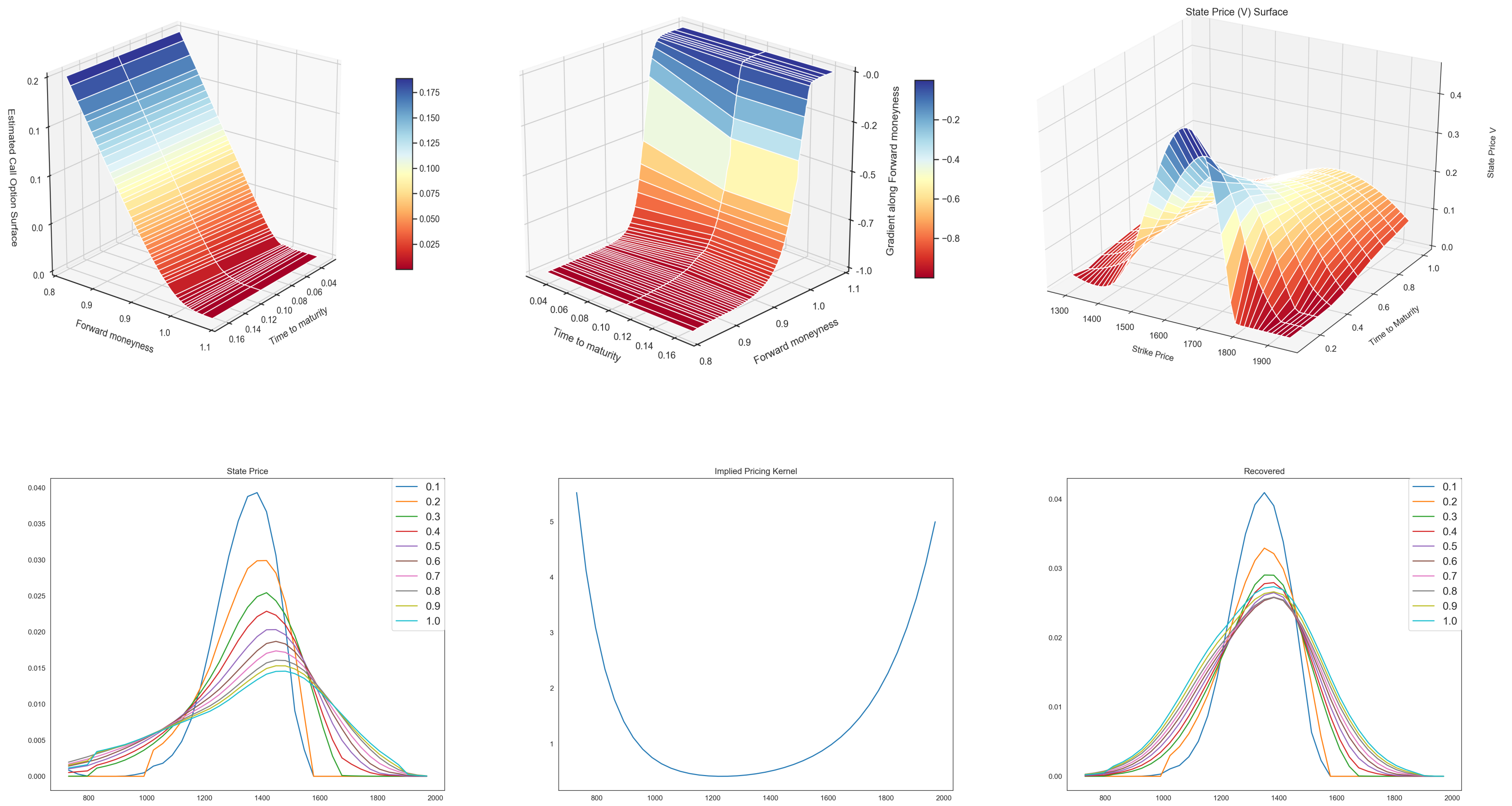
Motivated by these two problems and using a data-driven approach, in particular, I investigate the following questions:

- Can I estimate the RND using machine learning technique?, If the answer is yes, which technique could I use? (Chapter 3 and 5)
- How do I incorporate the no-arbitrage constraints in the machine learning framework? (Chapter 4 and 5)
- Compare to existing non-parametric method, whether my approach shows a better performance? (Chapter 5)
- Differ from previous discretization approach, if I examine TRT in a continuous time setting, what information do I recovered? (Chapter 6)

1.3 Findings and Contributions

Differing from many existing approaches that extract RND using specific parametric model, in this thesis, I try to solve the first problem from a machine learning perspective. In Chapter 5, I develop a data-driven estimator using a support vector machine(SVM). Including the bid-ask spread and no-arbitrage conditions, I modify the standard SVM into a linear programming framework. Compare with other methods, my approach yields four advantages. First, it is a fully nonparametric method and does not rely on the assumption of a prior

Figure 1.3.1: Key Results



Note: Figure 1.3.1 plots the key results in thesis. The top panel shows the extracted risk-neutral information: estimated call price surface, its first-order derivative and state price surface on 02/07/2013. The bottom panel displays the recovered results of Walden (2017)'s unbounded diffusion Recovery Theorem on 05/03/2008. The first two figures in top panel are plotted under forwards measure. In each subplot, selected time maturities (0.1 to 1.0) are indicated by different color.

distribution. Second, by setting trader's own error tolerance, it incorporates the information of bid-ask spread into an estimation framework. Third, it is a universal approach that incorporates all arbitrage-free constraints. Fourth, compared to the neural network, my approach could easily produce a smooth RND due to the explicit complexity control. The top panel of Figure 1.3.1 displays the estimation results of L_1-SVM on July 2, 2013. Consistent with option pricing theory, my state price density estimator yields a unimodal, smooth and positive surface.

To solve the second problem, I empirically implement Walden (2017)'s unbounded diffusion Recovery Theorem using S&P 500 index option. As shown in the bottom panel of Figure 1.3.1, I find that the recovered probability exhibits thinner left tails than its risk-neutral counterparts. Using Audrino et al. (2015) 's quadratic loss function in my L_1-SVM framework, I show that the quadratic loss function in state price estimation does not guarantee a well-recovered result. Finally, to further explain why The Recovery Theorem fails, I apply a bivariate unbounded recovery and confirm the existence of the martingale component in Ross's pricing kernel assumption.

To summarize, this thesis contributes to the previous studies in several three ways: First, I propose a new a data filter approach based on three principles: representative, accurate and no arbitrage in Chapter 4. I show that comparing with Zhang and Xiang (2008) and other studies, my filtered results contain less noise. Second, I contribute to the RND estimation studies by developing a more efficient linear programming support vector machine (L_1-SVM) estimator. Incorporating all no-arbitrage restrictions and bid-ask spread, in Chapter 5, I show that my method establishes a somewhat better accuracy and is universally applicable. Third, I provide the first empirical evidence for applying Walden (2017)'s unbounded diffusion recovery. Using the S&P 500 index option and synthetic data generated by Ornstein–Uhlenbeck (OU) process, I find consistent evidence on the recovered probability distributions with Jackwerth and Menner (2017). I show that the recovered probability is not the real-world probability.

1.4 Thesis Structure

This thesis contains seven chapters. Chapter 1 introduces the research background and gaps in existing studies. This chapter summarizes the main findings of the thesis. Chapter 2 gives a brief review of option theory, namely change of measure technique, Black-Scholes Model and no-arbitrage conditions for option price. In Chapter 3, I introduce two machine learning techniques and review their applications in finance. Chapter 4 shows the procedure for constructing option data panel and outlines the data filter rules. With the knowledge from proceeding chapters, I propose a new state price density estimator based on support vector machine (L_1-SVM) in Chapter 5. In this chapter, I extend the standard SVM to incorporate the no-arbitrage conditions and compare its performance with another four models. Chapter 6 investigates the problem of what information is recovered from TRT. I empirically implement Ross (2015) in the single diffusion process and bivariate state variables case. Chapter 7 concludes and suggests the future research direction.

Chapter 2

A Brief Overview of Option Theory

> People like the model because they can easily understand its assumptions. The model is often good as a first approximation, and if you can see the holes in the assumptions you can use the model in more sophisticated ways.
>
> —Black (1990)

This chapter presents the necessary mathematical finance foundation for this thesis. In section 2.1, I start by introducing two fundamental theorems in mathematical finance. These two theorems are prerequisites to link the no-arbitrage pricing principle with option pricing. Under no-arbitrage principle, I model the option price as a stochastic process and derive the Black-Scholes formula. In the last section, I summarize the sufficient and necessary conditions for incorporating no-arbitrage directly on a call option price surface.

2.1 Fundamental Theorem in Mathematical Finance

2.1.1 The First Fundamental Theorem of Asset Pricing

Although it is challenged by behavioral finance literature(e.g. Barberis and Thaler (2003)), the no-arbitrage principle is still the most important principle in financial modeling. The no-arbitrage principle arise from a basic belief about market behavior. Arbitrage is a trading strategy taking to make money when two or more securities are mis-priced relative to each other. When expressed in the mathematical finance literature, I connect the no-arbitrage principle with the probability measure. More precisely, if the market has no arbitrage opportunity, then there exists a unique equivalent martingale under which the underlying price process becomes a martingale.

Following Delbaen and Schachermayer (1994) and Dybvig and Ross (1989), let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. $\mathcal{F}$ is a right continuous information filtration , $\mathbb{Q}$ is the risk neutral probability and all possible outcomes are collected in Ω . S_t denotes a stochastic process [13] .

Definition 1. (Martingale) Let $S = S_t$ for $0 < t < \infty$ is a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$, if for any bounded stop time s (0<s<t<∞), the $E(S_t|\mathcal{F}) = E(S_s)$, then S is a martingale

Roughly, this means when conditioned on all the information I know today $\mathcal{F}$, My future expectation of a martingale process S is equal to its current value. When pricing the option, since the asset price has captured all the information in the market (EMH), the current expectation is equal to the asset price discounted by the risk-free rate.

$$E^{\mathbb{Q}}[\frac{S(T)}{M(T)}|\mathcal{F}] = \frac{S(t)}{M(t)} \tag{2.1}$$

The $\frac{S}{M}$ is a martingale process, where M is called market money account in the concept of numeraire.

Definition 2. (numeraire) A numeraire can be any asset that has a positive price process.

[13]The popular choices in mathematical finance literature are Wiener process and Levy process.

In particular, market money account, currency exchange rate and zero-coupon bond are popular numeraire choices in the mathematical finance literature.

Theorem 1. *(The First Fundamental Theorem of Asset Pricing) A market model in probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is no-arbitrage, if and only if there exists a probability measure that is equivalent to $\mathbb{P}$.*

Proof: please refer to Delbaen and Schachermayer (1998)

Essentially, the First Fundamental Theorem of Asset Pricing implies that under no-arbitrage principle, I could find a unique probability measure that enables the price process to have a constant expectation.

2.1.2 Change of Measure

It is clear that in real-world probability $\mathbb{P}$, the underlying price process is not a martingale. To benefit from the properties of a martingale and use the no-arbitrage principle, I need change the original probability measure $\mathbb{P}$ to another probability measure. This technique is called change of measure (unsurprisingly). For this, I first define the Radon-Nikodym Derivative.

Definition 3. (Radon-Nikodym Theorem) In probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a positive random variables Z and the probability measure $\mathbb{Q}$ is equivalent to $\mathbb{P}$, if

$$Z = \frac{d\mathbb{Q}}{d\mathbb{P}} \tag{2.2}$$

then Z is called the Radon-Nikodym derivative.

The Radon-Nikodym derivative can successfully change the measure when we know what we want to change into, but, in practice, the new probability measure is usually unknown. Therefore I consider Girsanov's Theorem.

Theorem 2. *(Girsanov's Theorem) Consider probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and let* $L(t)$ *is a adapted bounded process with* $L(0) = 1$ $\varphi(s)$ *is a positive process, if*

$$L(t) = exp(\int_0^t \varphi(s)dW^{\mathbb{P}}(t) - \frac{1}{2}\int_0^t \varphi 2(s)ds) \tag{2.3}$$

Where $W^{\mathbb{P}}(t)$ *is a standard Brownian Motion under* $\mathbb{P}$*. The process* $L(t)$ *also is the unique solution to*

$$dL(t) = \varphi(t)L(t)W^{\mathbb{P}}(t) \tag{2.4}$$

Then under a new probability $\mathbb{Q}$*,*

$$W^{\mathbb{Q}}(t) = W^{\mathbb{P}}(t) - \int_0^t \varphi(s)ds \tag{2.5}$$

$$dW^{\mathbb{Q}} = dW^{\mathbb{P}} - \varphi_t dt \tag{2.6}$$

Proof: please refer to Wilmott (2006)

The interesting aspect of theorem is that when changing to a new measure, the diffusion term of the SDE is unchanged while the drift term is changed.

Theorem 3. *(change of numeraire) If two different numeraire* N^1 *and* N^2 *and associated martingale measure* $\mathbb{Q}^{N^1}$ *and* $\mathbb{Q}^{N^2}$*, using Radon-Nikodym derivative, the following relationship holds:*

$$\frac{d\mathbb{Q}^{N^1}}{d\mathbb{Q}^{N^2}} = \frac{N^1(T)N^2(t)}{N^1(t)N^2(T)} \tag{2.7}$$

Proof. Since $\mathbb{Q}^{N^1}$ and $\mathbb{Q}^{N^2}$ are two martingale measures, according to definition 1, I obtain the following equations

$$E^{\mathbb{Q}^{N^1}}[\frac{S(T)}{N^1(T)}|\mathcal{F}] = \frac{S(t)}{N^1(t)} \tag{2.8}$$

$$E^{\mathbb{Q}^{N^2}}[\frac{S(T)}{N^2(T)}|\mathcal{F}] = \frac{S(t)}{N^2(t)} \tag{2.9}$$

Dividing these equations, I have

$$N^1(t)E^{\mathbb{Q}^{N^1}}[\frac{S(T)}{N^1(T)}|\mathcal{F}] = N^2(t)E^{\mathbb{Q}^{N^2}}[\frac{S(T)}{N^2(T)}|\mathcal{F}] \tag{2.10}$$

if I define $M(T) = \frac{S(T)}{N^1(T)}$,

$$\begin{aligned} E^{\mathbb{Q}^{N^1}}[M(T)|\mathcal{F}] =& \frac{N^2(t)}{N^1(t)}E^{\mathbb{Q}^{N^2}}[\frac{S(T)N^1(T)}{N^1(T)N^2(T)}|\mathcal{F}] \\ =& E^{\mathbb{Q}^{N^2}}[M(T)\frac{N^1(T)N^1(t)}{N^1(t)N^2(T)}|\mathcal{F}] \end{aligned} \tag{2.11}$$

Therefore, I can transform $\mathbb{Q}^{N^1}$ to $\mathbb{Q}^{N^2}$ using a Radon-Nikodym derivative defined by two numeraire assets. □

2.2 Option Pricing in a Nutshell

2.2.1 The Black-Scholes Model

Even though there has been some criticism of the Black-Scholes model, it is, without doubt, still the most widely used model in the derivative market. The Black-Scholes model provides a closed-form solution of the European option price and transforms the option pricing problem to solving partial differential equation(PDE).

If I make the following assumptions:

- The market has no-arbitrage.

- The market is frictionless: more precisely, there are no transaction costs on the underlying and the borrowing and lending interest rates are the same. All investors can immediately receive any available information.

- The underlying price follows a Geometric Brownian Motion(Top panel of Figure 2.2.1 shows an example of Geometric Brownian Motion prices paths)

$$dS(t) = rS(t)dt + \sigma S(t)dW^{\mathbb{P}}(t) \tag{2.12}$$

- The interest rate r is a function of t and volatility σ is constant.

- The underlying stock pays no dividends.

Let the $M(t)$ be market money account and $dM(t) = rM(t)dt$. According to definition 1, the risk-free discounted option price is a martingale, so

$$V(t,S) = M(t)E^{\mathbb{Q}}[\frac{V(T,S)}{M(T)}|\mathcal{F}] \tag{2.13}$$

Using Ito's Lemma, I get

$$dV = (\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt + \sigma S dW^{\mathbb{Q}} \tag{2.14}$$

If I define risk-free discounted option price as $\Pi \equiv \mathit{\Pi}(t,S)$

$$\mathit{\Pi}(t,S) = \frac{V(t,S)}{M(t)} = E^{\mathbb{Q}}[\frac{V(T,S)}{M(T)}|\mathcal{F}] \tag{2.15}$$

Since the discounted option price depends on underlying price and time t, the change of portfolio is defined as

$$d\Pi = d(\frac{V}{M}) = \frac{1}{M}dV - \frac{V}{M^2}dM = \frac{1}{M}dV - r\frac{V}{M}dt \tag{2.16}$$

Therefore

$$
\begin{aligned}
d\Pi = & \frac{1}{M}((\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2})dt + \sigma S dW^{\mathbb{Q}}) - r\frac{V}{M}dt \\
= & \frac{1}{M}((\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV)dt + \frac{\sigma S}{M}dW^{\mathbb{Q}}
\end{aligned} \tag{2.17}
$$

Since the risk-free discounted option is a martingale, therefore, it is drift-less. This gives us

$$
\frac{\partial V}{\partial t} + rS\frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0 \tag{2.18}
$$

The above equation is PDE of Black-Scholes. Consider the payoff condition of call option price

$$
V_C(T, S) = max(S(T) - K, 0) \tag{2.19}
$$

I get a unique solution for the PDE for call option price(see bottom panel of Figure 2.2.1 for an example of call price surface using Black-Scholes)

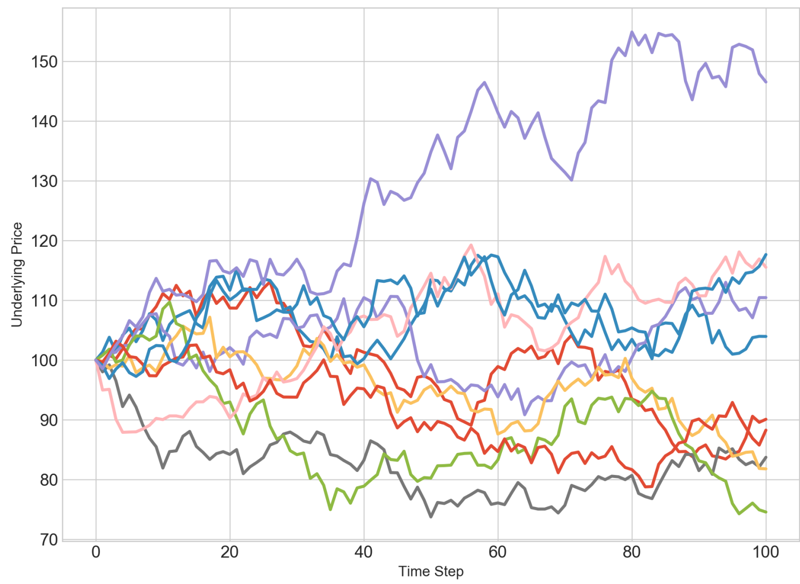
$$
\begin{aligned}
V_C(t, S) = & sN(d_1) - Ke^{-r\tau}N(d_2) \\
d_1 = & ln(\frac{S}{K}) + (r_\tau + \frac{1}{2}\sigma^2)\tau \\
d_2 = & d_1 - \sigma\sqrt{\tau}
\end{aligned} \tag{2.20}
$$

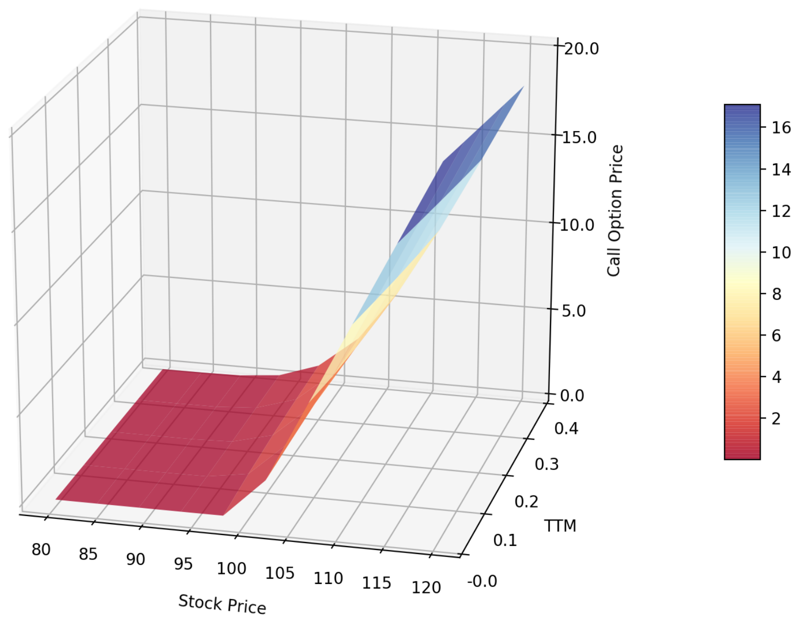
where $\tau = T - t$,$N(.)$ is cumulative distribution function of a standard normal distribution. Using put-call parity, I can get the European put option price

$$
V_C(t, S) = Ke^{-r\tau}N(-d_2) - SN(-d_1) \tag{2.21}
$$

Several studies have relaxed the assumptions of Black-Scholes including assuming stochastic volatility(e.g. Heston (1993)), allowing different diffusion pro-

Figure 2.2.1: Geometric Brownian Motion Price Paths and Black-Scholes Call Price Surface





Note: The top panel plots the 10 stock price paths which follows the Geometric Brownian Motion. The price is plot against time step. I divide the time t into 100 steps. The current stock price is 100. The bottom panel displays the call option price surface based on the Black-Scholes model. The interest rate r is 0.01, constant volatility is 0.1 and the strike price is 100.

cess such as Levy process, extending to the commodity market (e.g.Black (1976)) and considering dividend yield (e.g.Merton (1973)).

2.2.2 Numerical Techniques

The objective of option pricing is to determine the "fair value" of the option. Generally following the idea of Black-Scholes, I could model the underlying with a specified diffusion process and derive the PDE as in the previous section. Then I could get the analytical solutions for option price by applying stochastic calculus and solving the PDE. However, analytical solutions are not always possible and the case may be more complicated for the exotic options. As a consequence, four groups of numerical methods are often employed. All four methods are good at some perspective such as accuracy or easily converge but none of them is suitable and efficient for all problems. More detailed description of four numerical methods can be found in Wilmott (2006) and Hull and Basu (2016). I briefly summarize the advantages and disadvantages of my methods as following: first, Lattice(tree) method is the most straightforward and easily understood approach. Although Cox et al. (1979) prove that tree methods will converge to the correct option values, the convergence rate is relatively poor. The second approach is The Finite Difference, which is introduced by Brennan and Schwartz (1977) and is a direct approach to approximate the solution of the PDE. This approach provides reliable results for early exercise and low dimensional problems and performs poorly for path-dependent and high dimensional cases. Third, the Monte Carlo method, which simulates the underlying price path by random sampling and uses the average as option price. Although relatively slow, it yields excellent results for high dimensional and path dependent problems. Last but not the least is Numerical Integration via the quadrature method, which is an effective technique in both low and high dimensional problems. Mathematically, the idea behind the quadrature method(QUAD) is to approximate the areas under curves. In option pricing, Newton et al. (Andricopoulos et al. (2003, 2007); Chen et al. (2014)) present a series papers to apply QUAD in the classical Black-Scholes-Merton world, multi-asset and various underlying process.

2.2.3 Link to Equilibrium Pricing Model

As a matter of fact, the risk neutral pricing could connect to neoclassical finance with the no-arbitrage argument. According to Ross (2009), the no-arbitrage is equivalent to a positive linear pricing operator in asset pricing model.

Theorem 4. *(The Fundamental Theorem of Finance) The following three statements are equivalent:*

1. No Arbitrage(NA)

2.The existence of a positive linear pricing operator that prices all assets

3.The existence of a (finite) optimal demand for some agent who prefers more to less

Please refer to Dybvig and Ross (1989)for proof

The linear pricing operator is the necessary condition for setting up the equilibrium pricing model. The key idea of the equilibrium pricing model is that if the economy is characterized by a representative agent and no arbitrage, then the asset price is represented as a linear pricing operator multiplying future payoffs. The representative agent assumes that there exists a single agent in the economy who can represent the preferences of all investors. Strictly speaking, this is a strong and restrictive assumption. However, it implies a general result because it independent of an individual's initial wealth. The general equilibrium pricing model can be shown as

$$P_t = E_t[M_t Z_t] \tag{2.22}$$

where P_t is the asset price, Z_t is the future payoff and M_t is the linear pricing operator or stochastic discount factor(SDF). In the consumption-based capital asset pricing model developed by Lucas (1978), restricting the investor's consumption, the SDF is usually expressed as the marginal rate of substitution of consumption between different times. More specifically, $M_t = \beta\frac{u'(c_{t+1})}{u'(c_t)}$, c_{t+1} and c_t are the consumption at time t+1 and t, $u'(c_t)$ is marginal utility of consumption and β is the time preference discount factor. According to the model, I express the option price as expected value of stochastic discount factor weighted future payoffs. More formally, the option price at time t with payoff $Z(S_T)$ is

$$P_t = e^{-r\tau}E_t[M(S_T)Z(S_T)] = e^{-r\tau}\int_0^\infty M(S_T)Z(S_T)f(S_T)dS_T \tag{2.23}$$

Where S_T is a state variable and $f(S_T)$ is the payoff probability density under $\mathbb{P}$. To link the real-world probability $\mathbb{P}$ with the risk neutral density, I rewrite the Equation (2.23) as:

$$P_t = e^{-r\tau}E_t^*[Z(S_T)] = e^{-r\tau}\int_0^\infty Z(S_T)p^*(S_T)dS_T \tag{2.24}$$

Where $p^*(S_T)$ is the state price density or risk neutral density, which is real-world probability $\mathbb{P}$ discounted by stochastic discount factor. Compare Equation(2.24) with the option pricing equation under $\mathbb{Q}$ (see Chapter 5 for more detail), I can conclude that risk neutral density is a product of the real-world probability and stochastic discount factor.

2.3 No Arbitrage Conditions

Following Merton (1973) and Carr and Madan (2005), I classify the no-arbitrage conditions on option prices into two types. Type 1 focusses on arbitrage between option and risk-free bond. Type 2 emphasizes the arbitrage between different strikes and maturities. Denote $B(t,T)$ as the price of zero bond with $B(T,T)=1$, I summarize the no arbitrage conditions of option prices as follows

- Type 1: No arbitrage between option and risk-free bond

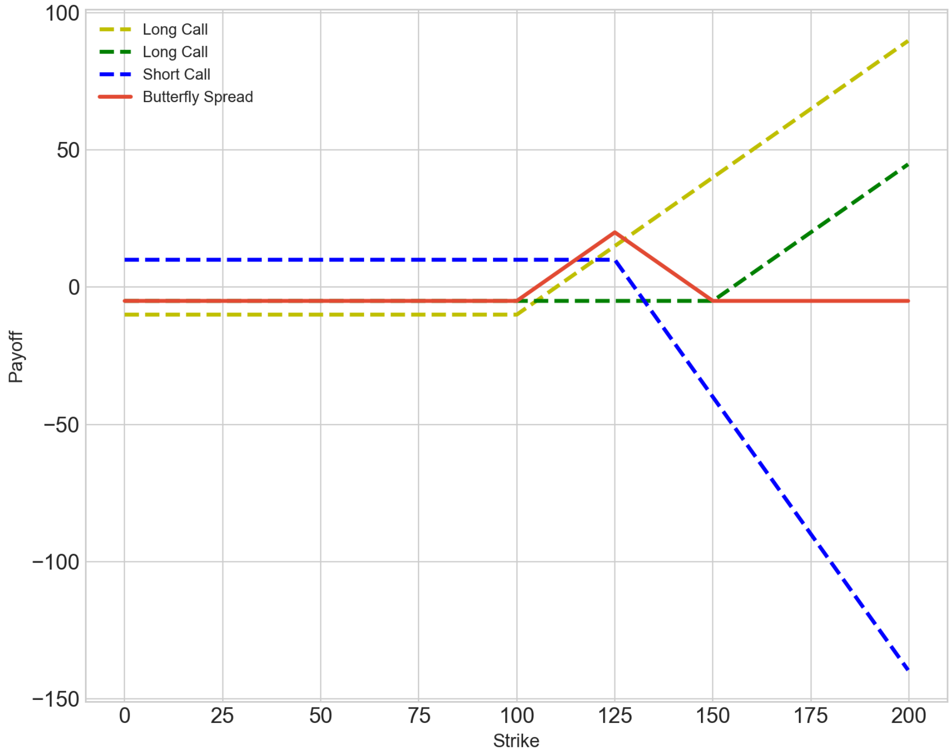
$$\begin{aligned} C(K,0) &= max[0, S_T - K] \\ [0, S_t - KB(t,T)] &\leq C(K,\tau) \leq S_t \\ C(K,\tau,S_t=0) &= 0 \\ C(K,\tau) + KB(t,T) &= P(K,\tau) + S_t \end{aligned} \tag{2.25}$$

The first equation defines the payoff of call option. The second and third equations restrict the call option. The last equation states put call parity. According to section 6 in Merton (1973), if I use Black-Scholes to price the European call option and the volatility is large than zero, then the Type1 no arbitrage constraints hold. Actually, as shown in section 2.2, the Black-Scholes equation is derived from Merton's definition of no-arbitrage.

- Type 2: No-arbitrage between different strikes and maturities

For the type 2 no arbitrage constraints, I use Carr and Madan (2005)'s idea of static arbitrage, which means if the call price surface is free from butterfly and calendar spreads then it is sufficient to define an arbitrage-free option price surface.

Figure 2.3.1: Payoff of Butterfly Spread



Note: This figure shows the payoff of a butterfly spread, with $K_1 = 100$,$K_2 = 125$ and $K_3 = 150$

- Absence of butterfly-spread arbitrage (Convexity respect to strike price): consider three call options $C_1(K_1)$,$C_2(K_2)$ and $C_3(K_3)$ with strike prices $K_1 < K_2 < K_3$. I can construct a butterfly spread whose payoff is always positive. For example, Figure 2.3.1 shows a butterfly spread with strike prices 100, 125 and 150.

$$B_f(K_1, K_2, K_3) = (K_3 - K_2)C(K_1) - (K_3 - K_1)C(K_2) + (K_2 - K_1)C(K_3) \quad (2.26)$$

Clearly, the payoff is positive. Arbitrage is possible if I could construct a butterfly spread with a nonpositive price. Therefore, I need to restrict the mapping such that $K \to C(K, \tau)$ is convex. Mathematically, this implies that when I twice differentiate the call price to strike price, the second derivative of call price is always greater than or equal to zero.

$$\frac{\partial^2 C}{\partial K^2} \geq 0 \tag{2.27}$$

- Absence of calendar-spread arbitrage (increasing respect to time to maturity): Consider two maturities $T_1 < T_2$ and with any strike $K^{T_1}$and K^{T_2} ,I define its forward value at maturity T as $F(T_1)$ and $F(T_2)$

$$\frac{K^{T_2}}{K^{T_1}} = \frac{F(T_2)}{F(T_1)} \tag{2.28}$$

If I construct a portfolio by buying a call option $C(K^{T_2}, T_2)$ and selling $\frac{D(T_2)F(T_2)}{D(T_1)F(T_1)}$ amount a call option $C(K^{T_1}, T_1)$, where $D(.)$ is the value of zero bond expiring at T . The $D(T, T) = 1$. At time T_1, if $S(T_1) > K^{T1}$, I define the payoff of call option $C(K^{T_1}, T_1)$ as

$$S(T_1) - \frac{F(T_1)}{F(T_2)} K^{T_2} \tag{2.29}$$

Otherwise, the payoff of call option $C(K^{T_1}, T_1)$ is 0. The value of portfolio for $S(T_1) > K^{T1}$ is

$$C(K^{T_2}, T_2) - \frac{D(T_2)F(T_2)}{D(T_1)F(T_1)}(S(T_1) - \frac{F(T_1)}{F(T_2)} K^{T_2}) \tag{2.30}$$

According to put-call parity, Equation (2.30) is equal to a put option with strike K^{T2} and maturity T_2. Since the value of portfolio for $S(T_1) \leq K^{T1}$ is $C(K^{T_2}, T_2)$, therefore in either case, the portfolio is always positive. Expressed differently, $C(K^{T_2}, T_1) > C(K^{T_1}, T_1)$ with $T_2 > T_1$ for any strike price. The call option price is increasing with time to maturity.

- Monotonicity respect to strike price:

$$-D(T) \leq \frac{\partial C}{\partial K}(K,T) \leq 0 \tag{2.31}$$

Consider I have two call option prices $C(K_1) < C(K_2)$, if I buy $C(K_1)$ and sell $C(K_2)$[14], at the time t, I receive $C(K_2) - C(K_1) > 0$. If the payoff is not monotonically decreasing with respect to the strike price, then at the maturity, $C(K_2) - C(K_1) > 0$. This provides an opportunity for arbitrage. The left-hand inequality means that a small increase in the strike will cause a risk-free discounted decrease in the call price. The proof can be found in Hull and Basu (2016).

- Price bounds: I denote the S is the current price of the underlying and $D(T)F(T) = e^{-\delta\tau}$

$$max(0, (D(T)F(T)S(T) - D(T)K)) \leq C(K,T) \leq SD(T)F(T) \tag{2.32}$$

Where $D(T)K$ is the present value of strike, $D(T)F(T)S$ is the underlying price-adjusted dividend.

This right-hand inequality means, when adjusted for the dividend, the value call option is less than holding an underlying . If the value option call option is great than underlying, then there is an opportunity for arbitrage by buying the stock and selling the call option. The left-hand inequality implies that the value call option is great than a forward at the same strike. Let us consider a portfolio which contains a call option with strike K and time to maturity T and K amount zero-coupon bond. When the strike price is greater than K, the value of portfolio is $S(T)$, otherwise, the value of portfolio is K. Hence, the value of portfolio is

[14] buy low and sell high

always greater than $S(T)$. Following the no-arbitrage principle and discounting the value to today, the zero-coupon bond is worth $D(T)K$ and $D(T)F(T)S(T)$ is stock price discounting for the dividend. In addition, the call price cannot be negative. Therefore, the left-hand inequality holds.

- The final value of the option price is its payoff $C(K, 0) = max(0, S - K)$

When the time to maturity is zero, I exercise the option right now, the value of option approaches its payoff. Since this condition is considered in type 1 constraints, please refer Merton(1973) and Hull and Basu (2016)for the detailed proof.

The above five conditions are summarized in The Theorem 2.1 in Roper (2010) from a mathematical martingale perspective(see Chapter 5 for more detail).

Chapter 3

Machine Learning in Finance

If you can't beat them, join them. Or maybe invest in them.

—Marsh (2018)

Machine learning has attracted a lot of attention over the past decade. From self-driving cars to speech/image recognition, from virtual assistants to cancer detection, it provides fruitful applications in many fields. In 2017, We witnessed an impressive breakthrough for artificial intelligence(AI)[15]. Google's AlphaGo, a machine learning program that plays the notoriously complex ancient board-game Go, quickly defeated top-ranked players. Training on millions of human professional games played in the past 3000 years, AlphaGo learned and mastered the game in the just 40 days[16]. The success of AlphaGo is a big leap forward for AI and assuredly it points towards the situation in the financial industry. Not surprisingly, investment banks, who have so much data for both buyers and sellers in the market, have already used machine learning to improve their performance. For example, JPMorgan[17] and Citi[18] use machine learning to make the competitive deals and develop trading strategies.

[15]Strictly speaking, machine learning is a subset of AI. Please see section 3.1 for more information

[16]see https://deepmind.com/blog/alphago-zero-learning-scratch/ for more detail.

[17]see https://www.bloomberg.com/news/articles/2018-08-20/inside-jpmorgan-a-bond-vet-builds-a-team-to-digitize-dealmaking for more detail

[18]see https://www.bloomberg.com/news/articles/2018-09-05/citi-s-credit-unit-starts-fintech-investment-group-under-zhang for more detail.

There is no doubt that finance will be changed by artificial intelligence. However, how to incorporate machine learning to solve the real financial problems and how machine learning impacts the structure of the financial industry is of great interest to both academic researchers and practitioners. In this chapter, I briefly review algorithms and applications of two of the most popular machine learning techniques in finance. In Section 3.1, I introduce the basic concepts of machine learning and provide an overview of the neural network and support vector machine approaches. Section 3.2 explores how these two techniques can be applied in solving financial problems such as stock market forecasts, sentiment analysis, and bankruptcy prediction. Focusing on the options markets, I review important studies that use machine learning.

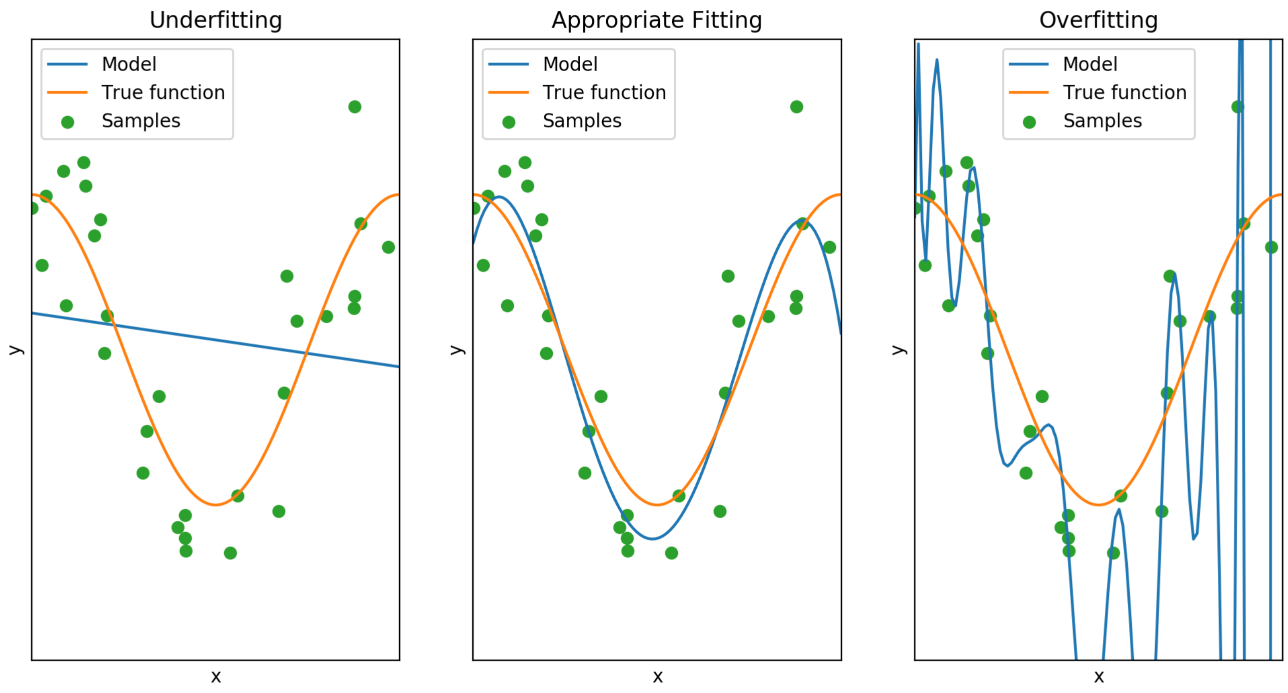
3.1 Machine Learning Overview

According to Stefanek (1987), AI is any computer system that exhibits elements of intelligence also shown in human behavior, such as understanding language, learning and reasoning. It includes broad approaches such as expert system, fuzzy system, neural network and support vector machine. Machine learning, as the most popular subset of AI, includes several statistical techniques that enable a machine to learn or complete a task. General speaking, "machine learning is using algorithms to extract information from input data and represent it in some type of model"(Patterson and Gibson (2017)). Based on the feature of training data, machine learning can be divided into two types: supervised and unsupervised. The objective of supervised learning is to find some type of model that maps inputs to the target value. Supervised learning is usually used in estimation and classification problems. In supervised learning, the training dataset clearly labels the input data and the target value. Unsupervised learning starts with no information about the target value and is a self-adjusted algorithm which trains itself to discover patterns in data. An example of unsupervised learning is the The Beatles-style song written by Flow Machine of Sony Computer Science Laboratory (CSL)[19].

According to Goodfellow et al. (2016), the machine learning framework generally contains four parts: the dataset, a loss function, an optimization algorithm and the estimated model. The dataset is usually split into training and testing subsets, with splitting done manually or via cross-validation. The loss function restricts model error and may be defined as the mean square error (MSE), cross-entropy loss or mean absolute error. The parameters of the estimated model are obtained from the training subset and the model's performance is evaluated using the testing subset. In practice, there is a trade-off between the quality of approximation and capability of generalization (Vapnik and Chervonenkis (2015)). That means, the learning algorithm may perform well in training but exhibit poor performance in prediction beyond the estimated dataset. From statistical learning theory, it can shown that a model cannot achieve a perfect fit to the input data while being flexible enough to predict out-the-sample data. Vapnik and Chervonenkis (2015) argue that training error and the model complexity yield a U-shape relationship. This problem is known as bias-variance trade-off or Vapnik-Chervonenkis (VC) dimension in statistical learning theory. Figure 3.1.1 shows an example of trade-off between model complexity and generalization using different polynomial orders. Compared with the true function, the left panel that uses a linear function shows a low learning ability but high generalization ability. The right panel demonstrates that the 20th-order polynomial estimator has

[19]see https://www.reuters.com/article/us-sony-algorithm-idUSKBN12H1ST for more information.

Figure 3.1.1: Model Complexity and Generalization



Note: This figure reports an example of the polynomial fitting. The input contains 30 data points and the true function is $f(x) = cos(2\Pi x)$. The left panel uses a first-order polynomial to estimate the problem. The middle panel uses the 4th-order polynomial as the estimated model and the right panel uses the 20th-order polynomial to approximate the true function.

impressive learning ability but poor generalization ability. Usually, the left panel is regarded as under-fitting and the right panel is called over-fitting. Apparently, finding the optimal complexity point of the learning algorithm is not easy because it is determined by the input dataset and pre-defined training error.

3.1.1 Neural Networks in a Nutshell

The neural network is an important technique in machine learning. Inspired by the human nervous system, McCulloch and Pitts (1943) first describes mathematically a simple neural network based on a threshold logic algorithm. This work exhibits a parallel estimating architecture and is considered a milestone in Artificial Neural Network(ANN). Slightly later, Rosenblatt (1958) conducts the first valid neural network implementation on an IBM computer. Defining a "perceptron" structure, Rosenblatt (1958) successfully classifies a pattern set in the input using a single-layer perceptron. Follow this research direction, Minsky and Papert (1960) empirical test the learning ability of a perceptron structure. They show that no matter how long it is trained, the perceptron structure is incapable of learning a non-linear function. This limitation is resolved by Rumelhart et al. (1986) who extend the previous 'single-layer' network to multi-layers. By introducing so-called hidden layers in network structure, they show that a neural network can be a universal approximator and learn any function. However, because of the difficulty of training multi-layer network, although neural network performs well in prediction and classification tasks, it became unpopular for several years. The age of the neural network actually begins again in last decade or so. Hinton et al. (2006) rebrand the neural network as deep learning.

A standard neural network consists of three elements: neurons, layers and connections between layers. Figure 3.1.2 illustrates a typical structure of a neural network. The input layer connects the input data x_i and the hidden layer. As shown in Figure 3.1.2, the neurons in the input layer receive input data and assign weights w_i plus a bias term b_i to them (gray lines in Figure 3.1.2). Taking these as input, the hidden layer that contains the transformation function (or activation function) maps a nonlinear transformation for weighted input. This layer is the most important and criticized part of the neural network. Since its values are not shown either in input and output data, mathematically, it is difficult to prove the resulting output in output layer is optimal. Obviously, in the prediction problem, the output layer only has one node but in the classification problem, the output layer contains many nodes (the number of nodes is equal to the number of classes). Figure 3.1.2 includes two hidden layers, therefore, the output from first hidden layer is further weighted by neurons and passed to the next hidden layer. The weights between different layers are updating repeatedly during the training process.

The neural network framework can be described as follows. Consider X $\{x_1, x_2, ...x_n\}$ as input vector, Y $\{y_1, y_{2,}......y_n\}$ as output vector. $f_1, f_2....f_l$ are the univariate transformation function for each hidden layer. The multilayer neural network can readily be defined as a chain structure $f(x) = f_l......(f_2(f_1))$. The length of the chain gives the depth of the neural network and the maximum number of neurons in hidden layers determines the width. If I express the problem in matrix form, the input data weighted by neurons defined as follows

$$\dot{Y} = X^T W + b \tag{3.1}$$

The three-layer neural network shown in Figure 3.1.2 is

$$\hat{Y} = f(X^T W + \bar{b})\tilde{W} + \tilde{b} \tag{3.2}$$

Where $\hat{Y}$ is estimated vector, X is the input vector. $\{W, \tilde{W}\}$ is a weight matrix and $\{b, \tilde{b}\}$ bias terms. f(.) is the transformation function (or activation function) in each hidden layer. The common choices of $f(.)$ are the polynomial function, radial basis function (RBF), spline function (see the next section for details) and sigmoid function[20]. The weight matrix is initialized with random values or prior information before training the neural network. The process of training the neural network is to solve the optimization problem

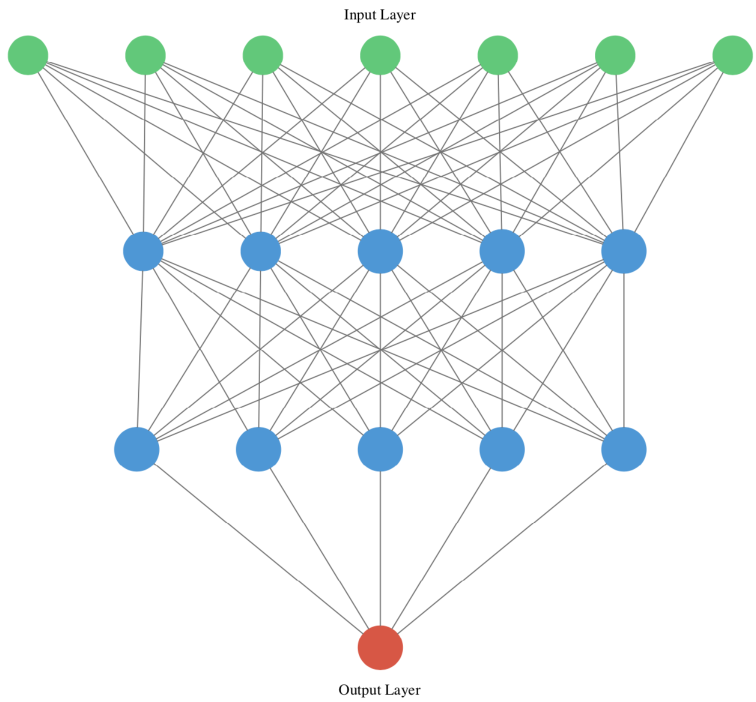
$$argmin_{\Theta}\frac{1}{N}l(Y,\hat{Y}) = argmin_{\Theta}\frac{1}{N}l(Y, f(X^T W + \bar{b})\tilde{W} + \tilde{b}) \tag{3.3}$$

Where $l(.)$ is a loss function that measures the deviation between the predicted value and target value. $\Theta = \{W, \tilde{W}, b, \tilde{b}\}$ is the parameters set.

Based on Equation(3.3), I argue that training the network is an unconstrained nonlinear minimization problem. The goal of training the network is to find the weight matrix in Equation(3.3) which gives minimal error measure such as MSE . This is usually achieved using back-propagation optimization developed by Werbos (1974). Back-propagation optimization is a gradient descent method with constant step size and the procedure can be summarized in the following way. First, with the initialized weight matrix, calculate the gradient of the loss function with respect to each weight, such as $\frac{\partial|Y-\hat{Y}|}{\partial w_i}$ for mean absolute error. Second, adjust the weight according to the calculated error. If the error is increasing with weights, reduce weights; otherwise increase the weights. This adjustment is

[20]These functions are reported in the Appendix.

Figure 3.1.2: Three-layer Neural Network Structure

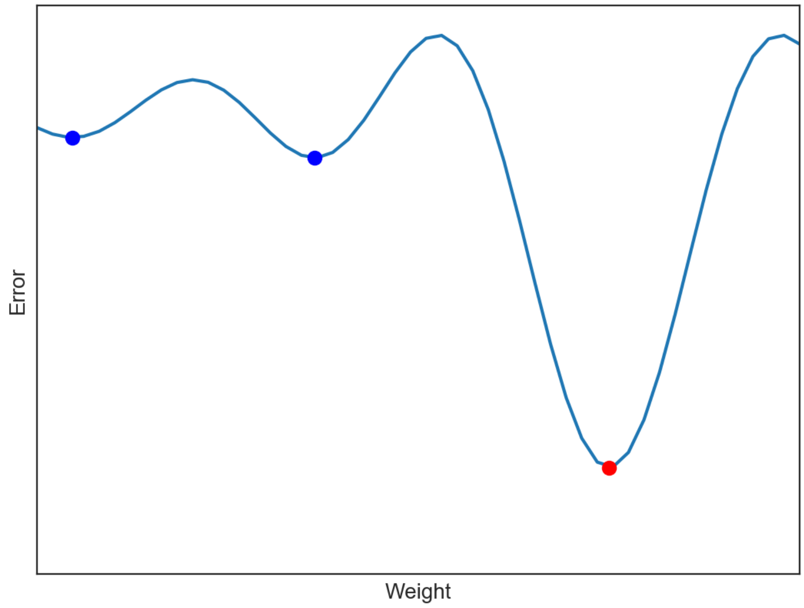


Note: This figure shows a structure of three-layer neural network. The green points are input layer, which contains 7 input features. The blue points represent the hidden layer that learning the input information. Gray lines denotes weights. The red point is output layer that gives the estimated value.

propagated backward; for instance, first correct the weights for the second hidden layer in Figure 3.1.2 and then the first hidden layer. After all weights are updated, the whole process starts over until the error meets the pre-determined threshold.

The neural network approach has been criticized for the stochastic weight initialization. The above optimization procedure generally leads to a minimum error. However, when the weight space exists and there are local and global minima, the optimal solution of the weight matrix is sensitive to the starting point. This effect is illustrated in Figure 3.1.3, where the blue points are local minima and the red point is global minimum. Simply, if the initial weight is between the first and second blue points, back-propagation moves the result to a local mini-

Figure 3.1.3: Weight Adjustment based on Back-propagation



Note: This figure shows the weight adjust. The x-axis is weight and the y-axis is the error between the target and the neural network predicted output. There are two local minimum and one global minimum in this search space. The blue points represent the local minimum and red point shows global minimum.

mum. If the initial weight is after the second blue point, the back-propagation finds the global minimum. Obviously, the global minimum is the preferred solution for Equation (3.3). However, with random initializing the starting weight matrix, there is no guarantee that guarantee the finding minimum is global. In particular, in the Figure 3.1.3, the probability of finding the minimum point is proportional to the distance between starting point to the relevant minimum points.

Another criticism of the neural network is that it is an ambiguous model. As shown in Figure 3.1.2, selecting the optimal model complexity, which is the degree in the polynomial estimated model, the number of knot for splines, is an essential issue in machine learning. For the neural network, the optimal model complexity could be controlled by choosing the width and depth of the neural network, which determines the number of connections in the network. For a given problem, a neural network with limited connections tends towards under-fitting while one with too many connections will over-fit. The optimal depth and width of a neural network are typically determined based on prior information or training dataset. Although Konishi and Kitagawa (2008)) try to identify the optimal number of connections using AIC (Akaike Information Criterion), Ding et al. (2013) argue that AIC shows inconsistent performance and tends to overfit the model. Consequently, I argue there is no general principle for finding the optimal model complexity of the neural network and the performance of the network is highly dependent on network design.

3.1.2 Support Vector Machines in a Nutshell

The other important technique in machine learning is the support vector machine (SVM). In contrast with the neural network, SVM always yields global minima and has explicit model complexity control. The SVM is also a universal approximator (Hammer and Gersmann (2003)) and the optimal model complexity can be restricted via regularization. The idea of support vector machines originates in statistical learning theory, which deals with drawing effective statistical estimation from small samples. The initial theoretical work of SVM is developed by Vapnik and Lerner (1963) and extended to regression by Vapnik et al. (1997). The key idea of SVM is the support vector. As shown in Figure 5.3.1, only these points that lie on the margin (dash line) define the estimated model. Therefore, the estimated model is determined by number and weights of support vectors.

Define a high dimensional feature space as $\Psi \in R^N$ and there exists a function

ϕ that maps a random input vector $X \in R^d$ from R^d to R^N. In this space, if I consider the estimated function as a class of linear functions

$$F_{n,\Psi} = \{f : f(X) = w^T\phi(X) + b\} \tag{3.4}$$

where $w \in R^d$ and $b \in R^d$ is weight and bias respectively.

Even if the estimated function is in high-dimensional feature space, I do not need to treat the high-dimensional problem directly. Replacing the inner product of two data points X_k, X_l with the corresponding kernel $K(X_k, X_l)$, I can transform the estimated problem into a low-dimensional feature space. This is known as "kernel trick" and guaranteed by Mercer's condition (Mercer (1909)).

Theorem 5. *(Mercer Condition) Let $K \in L^2(C)$ and $g \in L^2(C)$, where C is compact subset of R^d. The $K(t,z)$ describes an inner product of two feature space. If it is necessary and sufficient that*

$$\int\int_C K(t,z)g(t)g(z)dtdz \geq 0$$

where $t, z \in R^d$ and $g \in L^2(C)$, the continuous symmetric function K has an expansion

$$K(t,z) = \sum_{i=1}^{\infty} a_i\psi_i(t)\psi_i(z)$$

Proof: please refer to Mercer (1909)

Using the Mercer condition and defining $\phi(t) = \sqrt{a_i}\psi_i(t)$ and $\phi(z) = \sqrt{a_i}\psi_i(z)$, the kernel can be expressed as

$$K(t,z) = \phi(t)\phi_i(z) \tag{3.5}$$

Following Vapnik et al. (1997), the objective of SVM is to find a function that allows εdeviation from the target value (see Section 5.3 for more detail). Therefore, given the training data $\{(x_1, y_1), (x_2, y_2), ...(x_n, y_n)\}$, I minimize the empirical risk in the feature space

$$min_{\{w,b\}} \frac{1}{n} \sum_{i=1}^{n} |Y_i - w^T \phi(X_i) - b|_\varepsilon \tag{3.6}$$

where $|.|_\varepsilon$ is the ε insensitive loss function

$$|Y - F(X)|_\varepsilon = \{ \begin{array}{cc} 0 & |Y - F(X)| \leq \varepsilon \\ |Y - F(X)| - \varepsilon & otherwise \end{array} \tag{3.7}$$

To ensure the estimated function is as flat as possible, I seek a small w that satisfies the Equation(3.6). If I consider minimizing[21] the $||w||^2$, the optimization problem becomes

$$min \frac{1}{2} ||w||^2 \tag{3.8}$$

subject to

$$\{ \begin{array}{ll} Y_i - w^T \phi(X_i) - b & \leq \varepsilon \\ w^T \phi(X_i) + b - Y_i & \leq \varepsilon \end{array}$$

By introducing the slack variables ξ_i and ξ_i^* and using Lagrange multipliers, Equation(3.6) can be transformed into following dual problem[22]

$$max_{\{\alpha,\alpha^*\}} - \frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(X_i, X_j) - \varepsilon \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) + \sum_{i=1}^{n} Y_i (\alpha_i - \alpha_i^*) \tag{3.9}$$

[21] Please see Chapter 5 for the case of $|w|$

[22] I recommend reader to refer Schölkopf and Smola (2002) for detailed derivation.

subject to

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$$

$$\alpha_i, \alpha_i^* \in [0, c]$$

where α_i and α_i^* are optimal solution of Equation(3.9), which can be considered as Quadratic Programming (QP) problem. The dual representation of the estimated model is

$$F(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x, X_i) + b \tag{3.10}$$

The data points with non-zero α_i are considered as support vectors. $K(x, X_i)$ is the symmetric kernel that satisfies the Mercer's condition.

Although there are many programming packages such as quadprog in Matlab and Cvxopt in Python to solve the QP problem, Andersen et al. (2011) find that their performance varies with their algorithms. To get rid of the influence of different algorithm, in this thesis, I further simplify the optimization problem into a linear programming (LP) problem (see Section 5.3) and use Python Cvxopt.

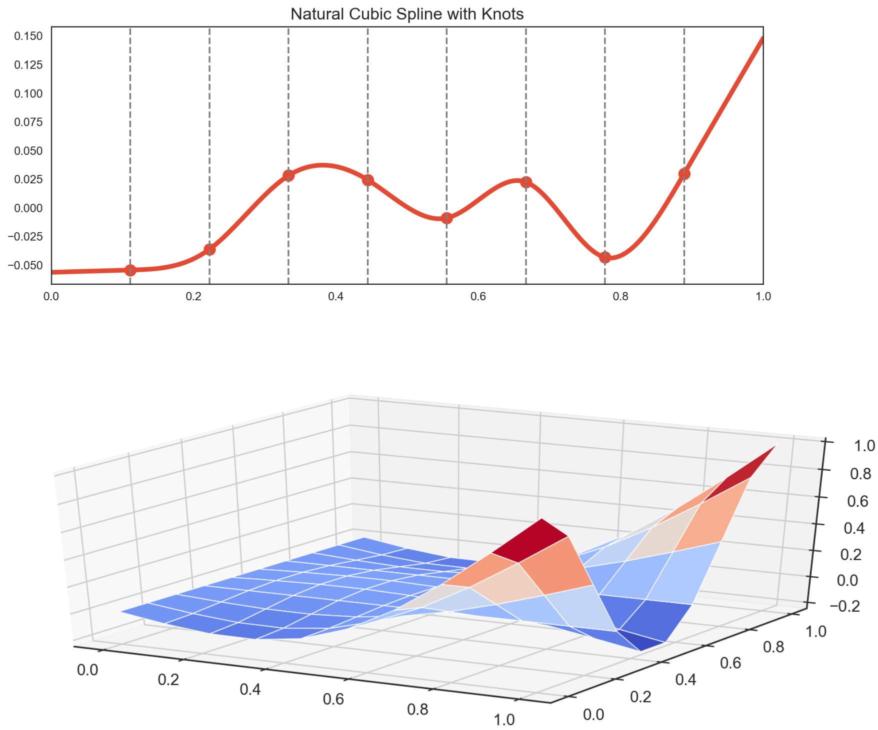
Since previous studies usually apply the spline function to approximate the call option price, following Birkhoff and de Boor (1965), I define the Spline and B-spline kernel as [23]:

Definition 4. (Spline) Consider a continuous function $s(x)$ with $x \in [a, b]$. If for each connection point $x_s \in [a, b]$ (or called knots), if a spline function's q-1 order derivatives exists, then this piecewise polynomial approximation $s(x)$ is called a spline function with order q.

The top panel of Figure 3.1.4 displays an example of a cubic spline which

[23] Another kernels are introduced in the Appendix

Figure 3.1.4: Spline Kernel



Note: The top panel shows an example of cubic spline kernel. The knot points are shown as red points. The bottom panel plots a B-spline(or tensor product spline) kernel.

has the form $s(x) = \alpha_s x^3 + \beta_s x^2 + \gamma_s x + \varepsilon$ in interval [0,1]. The second-order derivatives of $s(x)$ exist in each knot point (red points). If I further define this second derivative as equal to zero, then the spline function is called a natural spline. Extending the univariate spline to the bivariate case, I define the B-spline as follows.

Let the $U = \{u_{0,}u_1...u_h\}$ and $V = \{v_{0,}v_1...v_h\}$ be the knot vectors in u and v direction and the knot points are $p_{i,j}$, where $0 \leq i \leq m$ and $0 \leq j \leq n$, the B-spline kernel with degree p in the u direction and the degree q in the v direction is defined as the tensor product of two spline functions

$$k(u,v) = \sum_{i=0}^{m}\sum_{j=0}^{n} c_{i,j} K_{i,p}(u) K_{j,q}(v) \tag{3.11}$$

where $K_{i,p}(u)$ and $K_{j,q}(v)$ are basis spline functions of degree p and q, respectively. If I define the tensor product spline as matrices of $\mathbb{R}^{M_K \times N_K}$

$$Kmn = \begin{pmatrix} K(X_1)K(Y_1) & ... & K(X_1)K(Y_n) \\ \\ K(X_m)K(Y_n) & ... & K(X_m)K(Y_n) \end{pmatrix}$$

Then the matrix $\mathbb{K} \in \mathbb{R}^{(M_T \times M_K) \times (N_T \times N_K)}$ is simply defined as

$$K = (\begin{matrix} K_{11} & .. & K_{1N_T} \\ \vdots & \ddots & \vdots \\ K_{M_T 1} & ... & K_{M_T N_T} \end{matrix})$$

The bottom panel of Figure 3.1.4 shows an example of B-spline (or tensor product spline) kernel. As illustrated in the Figure 3.1.4, the B-spline kernel is a surface. Thus, compared with a univariate spline, it is a good way to avoid extrapolation in one direction when approximating a surface. Overall, following Bishop (2006), I compare the neural network and SVM in terms of ability in handling noisy data, processing large datasets, controlling model complexity, predictive accuracy and ease of operation in Table 3.1.1. Compared with a neural network, the support vector machine has high ability to handle noisy data, processing large data sest and controlling model complexity. This is because, as shown in right subplot in Figure 5.3.1 and suggested by Vapnik et al. (1997), the estimated function is only influenced by support vectors. Therefore, without considering every data point, the SVM can easily process a large dataset and control model complexity. Furthermore, since the outliers do not affect either the shape or slope of the

Table 3.1.1: Characteristics of Neural Network and SVM

	Neural Network	SVM
Ability to handle noisy data	poor	good
Process large data set	poor	good
Model complexity control	poor	good
Predictive accuracy	high	relative high
Ease of operation	difficult	easy

estimated function, the SVM has high ability to handle noisy data. However, since the SVM model is highly kernel-dependent, the predictive accuracy is very sensitive to the selection of kernel function. Therefore, compared with a neural network, the SVM has relatively lower predictive accuracy.

3.2 Machine Leaning in Finance

Although standard econometrics theory and parametric estimation methods dominated the finance literature for many years, motivated by the data-driven property of machine learning, various studies have applied machine learning in solving stock market forecasting, bankruptcy prediction and investor sentiment analysis. In this section, I provide a brief review focussing on the application of machine learning in time series forecasting and classification.

The stock market forecast problem can be regarded as a time series forecasting, which assumes that future outcomes are based (or partly based) on past observations. Forecasting time series using machine learning exhibits an attractive feature over standard econometrics tools since it is a data-driven approach. That said, even if the underlying relationship between input and output is difficult to specify, the machine learning approach can capture functional relationships. One of the successful applications of machine learning in the stock market is predicting stock volatility. GARCH-based volatility models stemming from Engle and Granger (1987)[24], have long served as a benchmark for modeling past return movement and future volatility in finance. However, with the influence of unexpected news, the GARCH family framework may infer incorrect return-volatility relationships and lead to wrong trading strategies [25]. Using the data-driven property and functional flexibility of machine learning, Donaldson and Kamstra (1997) and Pérez-cruz et al. (2003) use the neural network and support vector machine to estimate GARCH parameters respectively. They show that both techniques yield higher predicting ability than the common maximum likelihood estimated method. Following their studies, others investigate the performance of hybrid GARCH and machine learning models. For example, Cui et al. (2015) propose a new statistical arbitrage model based on a combination of a TGARCH model and a wavelet neural network. Using data from the Chinese metal futures market, they show that this hybrid model provides more accurate trading thresholds compared with the back-propagation neural network and simple TGARCH model.

The other application of machine learning in finance is classification. Essentially, the machine learning technique for classification is similar to logit regression, which labels the two possible dependent values as 0 and 1. However, thanks to its learning from examples, the machine learning technique can explore potentially "hidden" correlations among the predictive variables non-linear estimated function. Tam and Kiang (1992), estimate the bankruptcy problem and demonstrated the superiority of the neural network over discriminant analysis in terms of predictive accuracy and robustness. Later studies apply support vector machine (Chen et al. (2011)) and hybrid machine learning models (Chen (2014)) in the bankruptcy classification problem. In line with these studies and inspired

[24]Engle received the Nobel price in 2003 for his contribution of "analyzing economic time series with time varying volatility". Please see https://www.nobelprize.org/prizes/economics/2003/engle/facts/ for more detail

[25]This is called as asymmetry effect in literature.

by natural language processing, Luss and D'Aspremont (2015) and Pang et al. (2002) use machine learning to conduct sentiment analysis. Using newspaper articles and 10-K filings, they successfully estimate the relationship between investor sentiment and stock returns.

Studies applying machine learning in option markets started with Hutchinson et al. (1994). Using the S&P 500 index call option prices between 1987 and 1991, Hutchinson et al. (1994) apply two neural network models (the RBF network and multilayer perceptron) to pricing and hedging options. Considering the homogeneity of price and assuming constant volatility and the risk-free rate, they demonstrate that the neural network model outperforms the Black-Scholes (BS) formula. Subsequent studies further improve and extend this work by considering the stochastic volatility (Buchen and Kelly (1996)), including more input variables (e.g. interest rate (Qi and Maddala (1996)), investigating other markets (Bennell and Sutcliffe (2003)) and applying an alternative learning technique (Liang et al. (2009)).

From Equation(5.5), since risk neutral density is equal to the second derivative of the call option pricing function, the performance of estimation of state price density highly depends on the estimated option pricing model. In this thesis, I use the support vector machine to extract state price density (or risk neutral density) for two reasons. First, as noted in Section 3.1, the neural network has ambiguous model complexity control. The width and depth of the neural network, which determines the number of connections in the network, is selected from experience. Therefore, compared with SVM, the estimation model using a neural network may lead an over-fitting or under-fitting of the result. Second, based on the intuition of SVM (please see Chapter 5), the ε insensitive loss function in SVM naturally considers the bid-ask spread in the estimation model. However, as shown Table

3.1.1, the support vector machine (SVM) technique is easily affected by noisy data; therefore I apply a relatively comprehensive data filter in Chapter 4.

Chapter 4

Option Data Panel

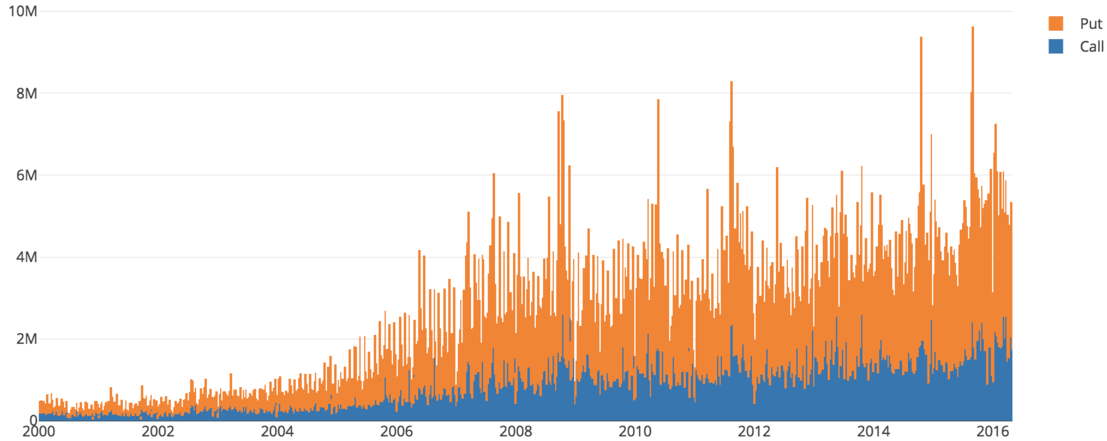
The empirical analysis of this thesis is based on a large daily option panel over 16 years. In section 4.1, I describe the dataset, including general characteristics and the reason of choosing S&P 500. Section 4.2 provide a detail description of data filter rules. Based representative, accurate and arbitrage-free principles, I develop a three levels of filter(data accuracy, liquidity and no arbitrage). Finally, in section 4.3, I show how to change the estimated framework from risk neutral measure $\mathbb{Q}$ to forward measure $\mathbb{Q}^{\mathbb{T}}$ and derive the no-arbitrage conditions for call price under forward measure.

4.1 Data Overview

This thesis examines a large daily cross-sectional option panel from January 5, 2000 to April 30, 2016. These were the latest available data when chapter 5 was the written. I investigate closing Standard & Poor's 500 (S&P 500) option price from OpionMetrics via Wharton Research Data Services(WRDS). The S&P 500 index option (option symbol SPX) is traded in the Chicago Board of Trade's Option Exchange (CBOE) between 8:30 to 15:15 CST in every trading day. An option on the S&P 500 is cash settled[26] in the morning on the third Friday of

[26] Option holder/writer pay the difference between strike and underlying price when the option expires

Figure 4.1.1: Daily Trading Volume of S&P 500 Index Option in my Sample



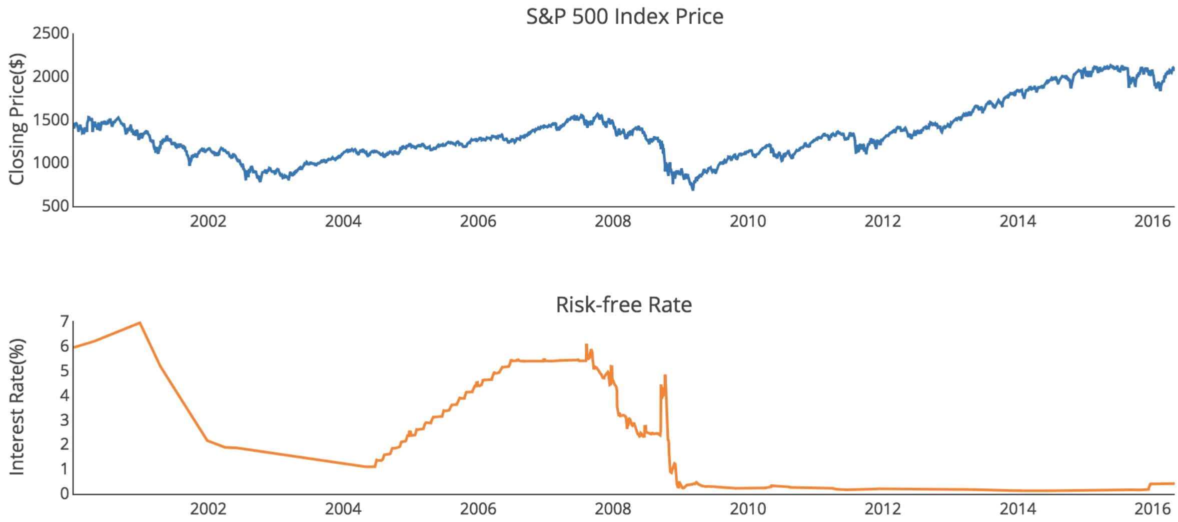
Note: This figure plots the daily trading volume of S&P 500 index option between January 5, 2000 to April 30, 2016

every month[27]. There are up 12 maturities available for each trading date and the strike interval is \$5. The underlying S&P 500 index is a value-weighted index that contains top 500 companies according to market capitalization. It is usually treated as a leading benchmark of the American equity market. In the sample period, there are 915 firms in the S&P 500 index because the constructed firms are updated over time as market capitalization changed. The companies list is in Appendix. The S&P 500 index price, trading volume, risk-free rate are also downloaded from OptionMetrics.

Figure 4.1.1 depicts the trading volume of S&P 500 index options in my sample. It is evident that put options are more active than call options. The trading volume of put options has several peaks during the sample period. For example, one peak point is near the 2008 financial crisis; I suggest that this

[27]See http://www.cboe.com/products/stock-index-options-spx-rut-msci-ftse/s-p-500-index-options for more details

Figure 4.1.2: Time Series of S&P 500 Index and Risk-free Rate



Note: This figure reports the monthly average of S&P 500 index and risk-free rate over the period January 5, 2000 to April 30, 2016. Both data are from OptionMetrics. The average dividend rate is 2.53% and the interest rate is 1.17% in this period.

overtrading of put options is caused by investor sentiment. Investors rushed to purchase put options so as to lock in current profits and protect themselves from the future market crash.

In Figure 4.1.2, I plot the monthly average of the S&P 500 index and the risk-free rate from 2000 to 2016. The S&P 500 index price dropped significantly in the 2008 financial crisis and the risk-free rate maintained a near zero value over the period from 2008 to 2016. The corresponding monthly average call price is plotted in Figure 5.4.1. As the S&P 500 includes a broad range of industries such as technology, retail, and financial services, its price is popular in previous literature as an indicator of consensus among market participants. Therefore, the corresponding S&P 500 index option reflects most investors' expectations about future market movements and the average attitude of market participants. That said, economically, investors in the S&P 500 index comprise a good indicator of the representative agent.

4.2 Data Filter

In this section, I present the data filters rule and process. The option price selected in this thesis is inspired by three principles: representative, accurate and arbitrage-free. To infer the general properties of call option prices, a representative dataset should be estimated. I construct my dataset with daily European option written on the S&P 500 index as an economically representative market portfolio. I use only the out of the money (OTM) options because these have higher liquidity and therefore the quoted price are closer to theoretical prices. The OTM option is defined as $K < F_0$[28] for put option and $K > F_0$ for call option. To increase the data quality and assuming put-call parity holds

$$C - P = e^{-\tau t}(F - K) \tag{4.1}$$

Taking advantage of this relationship, I need to get the daily interest rate as close as possible to market to transform the OTM put prices to ITM call prices. There are two ways to deal with the issue.

1. Interpolation

The interest rate is intuitively determined by the interest rate yield curve; therefore, I can linear interpolate the curve to get the daily interest rate. For example, if the interest rate for 1-month maturity is 0.97%, 3-month maturity is

[28] F_0 is at the money(ATM) forward

0.95%, the interest rate for maturity at 24 days is computed as:

$$\begin{aligned} r &= r_{1mth} - (r_{3mth} - r_{1mth}) * \frac{30-24}{61} \\ &= 0.97\% - (0.95\% - 0.97\%) * \frac{30-24}{61} = 0.9720\% \end{aligned} \tag{4.2}$$

2. Simple Regression

If I use the put call parity relationship and consider a linear regression of at least 4 put-call option pairs, the put call parity of Equation (4.1) can be expressed as:

$$C_i - P_i = \alpha + \beta K_i \tag{4.3}$$

Where $\alpha = Fe^{-r\tau}$, $\beta = e^{-rt}$ and interest rate $r = \frac{-ln(\beta)}{\tau}$, dividend yield is $\frac{-ln(\frac{\alpha}{F})}{\tau}$. This approach enables us to back out the daily interest rate and dividend at the same time.

To ensure my dataset is close to the market, in this thesis, I adopt both approaches. I first linearly interpolate the yield curve to get daily interest rate because I can get zero coupon rate from OptionMetrics database. I use this as my interest rate. Subsequently, I use Equation (4.3) to compute the daily dividend. Last but not least, I modify the data filters of Constantinides et al. (2013) and apply three levels of filter designed to obtain accurate and arbitrage-free S&P 500 option prices.

1. Data Accuracy Filter

- Identical observations: I drop all duplicated observations. When observations have the same identical terms (start date, expiration date, strike, option type) but differ in price, I keep the quotes whose implied volatility is significantly away from its moneyness neighbors.

2. Liquidity Filter

- Zero bid price : I remove quotes with zero bid prices to avoid illiquid options.
- Zero volume: in the same spirit, I remove quotes with zero volume.
- Days to maturity <7 or >365 : I remove data with maturity less than 7 days or more than 1 year. According to Constantinides et al. (2013), quotes with shorter maturity will tend to move erratically and quotes with longer maturity lack volume. I modify this filter with longer durations[29] because, from Figure 4.2.1, S&P 500 options are still active before 1 year.
- Moneyness<0.8 or >1.2: Constantinides et al. (2013) argue that option quotes in this range are thinly traded. As shown in Figure 4.2.1, I also observe this feature in my dataset. Therefore I remove option quotes with moneyness (the ratio of strike price to index price) below 0.8 or above 1.2.
- Implied volatility <5% or >100%: I remove all quotes with implied volatility lower than 5% or higher than 100% because as suggested by Constantinides et al. (2013) , option quotes in this range can be considered as illiquidity.

3. No arbitrage Filter

- Negative Implied interest rate: I remove quotes with negative put-call parity implied interest rate. For each available date and maturity, I use put call pairs with at least six strike prices to calculate implied interest rate[30].
- Negative Time Value: I discard quotes that have negative time value. Option prices consist of two components: intrinsic value and time value. Option quotes with negative time value show little information about investor's

[29] Constantinides et al.(2013) choose 0.5 years as days to maturity upper bound.
[30] I use put call parity to get implied interest that minimize implied forward pricing errors.

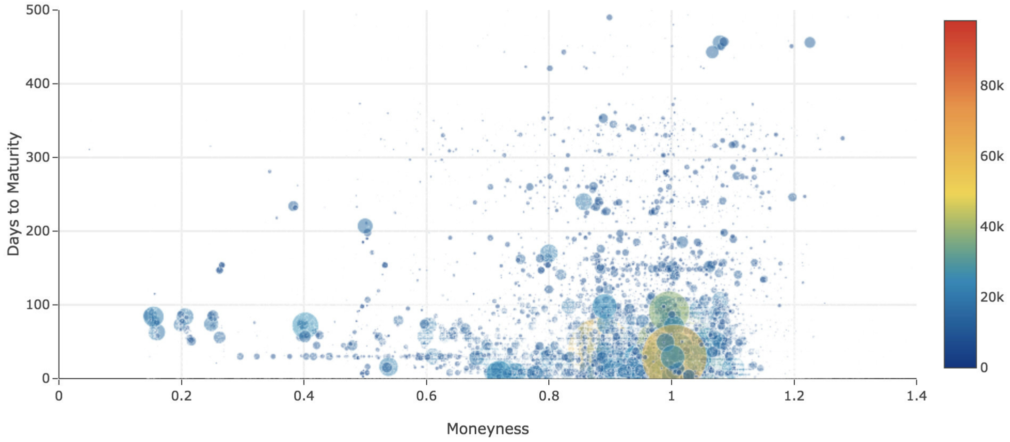
expectation because time value representing the amount of risk premium that investor is willing to pay.

- IV filter: I remove quotes whose implied volatility is one standard deviation away from the average among peers. I define peer group by the different levels of moneyness. More precisely, for each date and maturity, I fit the log implied volatiles in entire sample via a quadratic curve (separately for call and put options). I compute the relative distance of all observed IV from fitted IV and then I truncate the fitted curve to bins of moneyness with a width of 0.05(0.8, 0.85, ..., 1.2). After calculating the standard deviation for each moneyness bin, I discard quotes whose observed IV is one standard deviation apart from the fitted IV.

- Put-call parity filter: I remove any quote whose the put-call parity implied interest rate is more than one standard deviation away from the average among the peers. Peer group is defined as quotes with the same (date, time to maturity) pairs. I trim outliers in a similar way as with the IV filter. Specifically, I use the whole sample of distances of the put-call parity implied interest rates from the corresponding daily median implied interest rate to find the standard deviation of the corresponding distances.

Table 4.2.1 records the number of observations at each filtering level that are removed. Before the filters, I have a total of 8,261,170 observations. Level 1 filters remove 10 observations. The zero volume criteria in the Level 2 filters eliminate the most observations (5,433,167) and level 3 filters eliminate 8.3% of observations. The final number of observations in this thesis is 498,778, which is reasonable for three reasons. First, compared with Chiang et al. (2016),who also follows Constantinides et al. (2013) and gets total 404,822 observations on S&P

Figure 4.2.1: Trading Volume of S&P 500 on 2016



Note: This figure shows the trading volume of S&P 500 options. The data period is from 01 January 2016 to 30 April 2016.

500 options between 1996 and 2011, my dataset is relative large. Second, the No Arbitrage assumption is a strong assumption about option price observations in this thesis because the real traded option prices are determined in the real market, in which there are arbitrage opportunities. Third, as discussed in section 4.2, four No Arbitrage (NA) filters are supported by option pricing theory. Figure 4.2.1 plots the trading volume of the first 4-month in 2016 and the number of trading volume is indicated by color. Based on the liquidity filter, I remove the option whose days to maturity are greater than 365. Also, the trading volume for options with moneyness before 0.8 and days to maturity greater than 100 are scarce and I select the options between 0.8 and 1.2 moneyness and option that expire after 1 year.

I consider a sample of option price data on 27 August 2008 to illustrate the filter process and check the no-arbitrage property of the processed result. The

Table 4.2.1: Number of Observations

		Deleted	Remaining
Starting	Calls		4,130,624
	Puts		4,130,546
	All		8,261,170
Accuracy Filters	Identical Terms	10	
	All		8,261,160
Liquidity Filters	Zero Bid /Ask	708,202	
	Zero Volume	5,433,167	
	Days to Expiration Boundaries	262,388	
	Moneyness Boundaries	254,545	
	IV Boundaries	417,832	
	All		1,185,026
NA validation	Negative Implied Interest Rate	289,542	
	Negative Time Value	232	
	IV Outlier	173,725	
	Implied Interest Rate Outlier	222,749	
	All		498,778
OTM Options			390,320

Note: Table 4.2.1 presents the number of observations after each filter. The sample period is from 01/01/2000 to 04/30/2016. The moneyness is defined as the ratio of strike price to index price. IV is implied volatility, NA is no arbitrage and OTM is out the money. Implied interest rate is calculated based on linear regression(Equation 4.3). The time value of option is defined in Equation(4.4) and (4.5).

raw option data and calculated call price are shown in Table 4.2.2. The forward price is \$1465.40 and the S&P 500 index price is \$1281.66, which means that the S&P 500 index dividend rate is higher than the risk-free interest rate. This is confirmed by statistics of my full dataset. The average dividend rate is 2.53% and the interest rate is 1.17%. Figure 4.2.2 displays the distribution of call prices after filters. I note that there are three maturities available on this trading date. Clearly, the options trading of deep OTM is very thin even for S&P 500 index options[31]. Although the dataset does not have a large coverage of call prices as high-frequency data, I argue my dataset is efficient enough to cover the call price surface. Compare with Chiang et al. (2016), who gets total 404,822 observations, my dataset is relative large. Also as shown in Figure 5.4.1, my dataset is large enough to plot a well-defined call price surface.

Furthermore, to compare my no arbitrage filter with Zhang and Xiang (2008), I plot the time value of call option price on 04/11/2003 in Figure 4.2.3 and check whether the dataset shows no arbitrage. The time value of an option is defined as

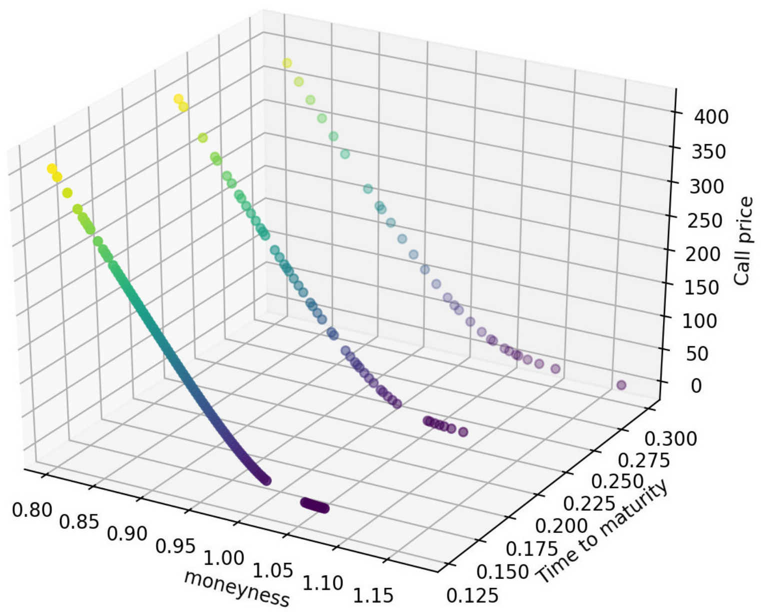
$$c_{time}(k,\tau) = c(k,\tau) - e^{-rt} max(F_0 - K, 0) \tag{4.4}$$

$$p_{time}(k,\tau) = p(k,\tau) - e^{-rt} max(K - F_0, 0) \tag{4.5}$$

If the put call parity relationship in Equation (4.1) holds, this means that time value of call price $c_{time}(k,\tau)$ and put price $p_{time}(k,\tau)$ are equal. Compare Figure 4.2.3 with Figure 2 in Zhang and Xiang (2008), I argue that my dataset is arbitrage free and the time value of call and put prices almost coincide. The

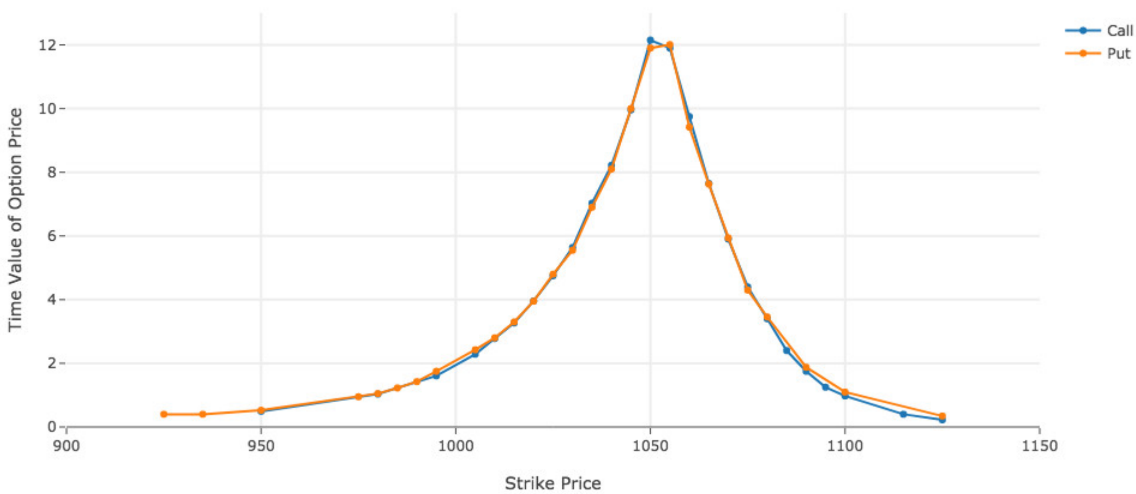
[31]The deep of OTM option means the strike price is far away from current index price. Consequently, buying a deep OTM option is expecting that a extreme increase of index price.

Figure 4.2.2: Distribution of Call Option Price after Filters on August 27, 2008



Note: This figure shows the S&P 500 index call option price across different strike and time to maturity on 27/08/2008. The S&P 500 index price is $1281.66.

Figure 4.2.3: Time Value of S&P 500 Option Price on November 04, 2003



Note: This figure shows the time value of S&P 500 index call and put options on 04/11/2003. The days to maturity is 17 and the S&P 500 index price is $1053.25.

Table 4.2.2: Option Data Panel on August 27, 2008

Options	Strike	Bid	Ask	Volume	Call Option Price
	1435	20	22.2	3177	51.40
	1440	20.9	23.5	7168	47.52
	1445	22.4	25	5322	44.03
Put	1450	24.2	25.3	25947	40.10
	1455	25.6	28.2	2594	37.26
	1460	27.5	29.9	3339	34.08
	1465	29.2	31.8	10440	30.89
ATM	1465.40				
	1470	26.5	29.1	142	27.80
	1475	25	25.8	4931	25.40
	1480	22	23.5	2574	22.75
Call	1485	18.5	20.9	355	19.70
	1490	16.1	18.5	888	17.30
	1495	13.9	16.3	265	15.10
	1500	12.4	13.5	29721	12.95

Note: Table 4.2.2 shows the market data on 27/08/2007. The days to maturity is 24. The computed ATM forward price is \$1465.40 and S&P 500 index price is \$1281.66. To give reader a general idea, this table only presents a part of data.

difference in peak time value of call price in Zhang and Xiang (2008) is significantly higher than mine. That said, my dataset contains less noise than Zhang and Xiang (2008).

4.3 Change of Measure for Estimated Framework

4.3.1 Change to Forward Measure

To compare my L_1-SVM estimator with other models in Chapter 5, I investigate the estimated call price problem under the forward measure (from $\mathbb{Q}$ to $\mathbb{Q}^{\mathbb{T}}$). Define $F(t,T)$ as the forward measure and $F(T,T) = 1$. According to equation(2.11),

$$E^{\mathbb{Q}}[V(T)|\mathcal{F}] = E^{\mathbb{Q}^{\mathbb{T}}}[V(T)\frac{M(t)P(T,T)}{M(T)P(t,T)}|\mathcal{F}] \tag{4.6}$$

$$\frac{d\mathbb{Q}^{\mathbb{T}}}{d\mathbb{Q}} = \frac{e^{\int_t^T r(t)dt}}{F(t,T)} \tag{4.7}$$

Previous studies have different assumptions regarding interest rate and dividends. Some studies allow constant interest rate and some for a deterministic rate (see Chapter 5 Table 5.5.1 for summary). To compare existing estimators with the proposed estimator, I normalize the call price to eliminate the influence of interest rate and dividend. Recall Merton(1973)'s model, which includes dividends in the Black-Scholes framework. I define forward price $F(t,T) = se^{(r-\delta)\tau}$ and simplify the $e^{\int_t^T r(t)dt}$ as e^{rt}.

$$C(K,\tau) = se^{-\delta\tau}N(d_1) - Ke^{-r\tau}N(d_2) \tag{4.8}$$

$$\frac{C(K,\tau)e^{rt}}{se^{(r-\delta)\tau}} = \frac{se^{-\delta\tau}N(d_1)e^{rt}}{se^{(r-\delta)\tau}} - \frac{e^{rt}Ke^{-r\tau}N(d_2)}{se^{(r-\delta)\tau}} \tag{4.9}$$

$$\frac{C(K,\tau)}{se^{-\delta\tau}} = N(d_1) - \frac{Ke^{-r\tau}N(d_2)}{se^{-\delta\tau}} \tag{4.10}$$

If I define the forward moneyness $k = \frac{K}{F(t,T)} = \frac{K}{se^{(r-\delta)\tau}}$, then the Black-Scholes-Merton formula can transform to

$$\frac{C(K,\tau)e^{rt}}{F(t,T)} = N(d_1) - kN(d_2) \tag{4.11}$$

$$d_1 = \frac{ln(\frac{s}{K}) + (r_{t,\tau} - \delta_\tau + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} = \frac{ln(\frac{1}{k}) + \frac{1}{2}\sigma^2\tau}{\sigma\sqrt{\tau}} \tag{4.12}$$

$$d_2 = d_1 - \sigma\sqrt{\tau} \tag{4.13}$$

Compare Equation(4.8) with Equation(4.11), the Equation(4.11) is equivalent to a European call option price with underlying is 1, interest rate r and dividend yield δ equal to 0. Hence, by transforming the estimating framework to forward measure, I eliminate the influence the interest rate and dividend yield.

4.3.2 No Arbitrage Conditions of Call Price under Forward Measure

There are two ways to restrict the arbitrage free call price. One is from the state price density perspective and the other is from the option strategies perspective (see Chapter 2 section 2.3). In this section, I show how to get the similar no arbitrage conditions of Roper (2010) under the forward measure. First, recall the no arbitrage conditions under the risk neutral measure $\mathbb{Q}$. The call price is the numerical integration of the payoff and risk neutral density (RND).

$$C(S_t, K, \tau, r, \delta) = e^{-r\tau} E_t^Q[max(S_T - K, 0)] = e^{-r\tau} \int_0^\infty \max(S_T - K, 0) q^*(S_T) dS_T \tag{4.14}$$

Following Equation(4.7) and define $k = \frac{K}{F(t,T)}$

$$C(S_t, k, \tau, r, \delta) = E_t^{Q^T}[max(\frac{S_T}{F(t,T)} - \frac{K}{F(t,T)}, 0)] = \int_0^\infty \max(\frac{S_T}{F(t,T)} - k, 0) q^{Q^T}(S_T) dS_T \tag{4.15}$$

Take the first order derivative

$$\frac{\partial C(k,\tau)}{\partial k} = -\int_0^\infty q^{Q^T}(S_T) dS_T = -(1 - F(k)) \tag{4.16}$$

Where $F(k)$ is cumulative distribution function of the transition probability q^{Q^T} under forward measure

$$\Rightarrow \frac{\partial C(k,\tau)}{\partial k} = (F(k) - 1) \tag{4.17}$$

As $F(k)$ is always greater than zero:

$$\Rightarrow F(k) \geq 0$$

$$\Rightarrow \frac{\partial C(k,\tau)}{\partial k} \geq -1 \tag{4.18}$$

Since q^{Q^T} is transition probability, based on a generic property of probability density, it is non-negative.

$$\Rightarrow q^{Q^T}(S_T) \geq 0 \tag{4.19}$$

Based on Equation (4.16),

$$\Rightarrow \frac{\partial C(k,\tau)}{\partial k} \leq 0$$

Therefore, recalling Equation (4.18), I get

$$-1 \leq \frac{\partial C(k,\tau)}{\partial k} \leq 0 \tag{4.20}$$

Further differentiate the Equation (4.16),

$$\frac{\partial C(k,\tau)}{\partial k^2} = q^{Q^T}(S^T) \geq 0 \tag{4.21}$$

Since $C(S_t,k,\tau,r,\delta) = E_t^{Q^T}[max(\frac{S_T}{F(t,T)} - \frac{K}{F(t,T)}, 0)]$, by Jensen's inequality(Jensen

(1906)), the max function is convex, thus

$$C(S_t, k, \tau, r, \delta) \geq max(E_t^{Q^T}[(\frac{S_T}{F(t,T)} - k, 0)]) = max(1-k, 0) \tag{4.22}$$

This result can also be obtained by setting $D(T)F(T) = 1$ and $S(T) = 1$ in Equation(2.27) and (2.32). When $k = 0$,$C(S_t, k = 0, \tau, r, \delta) = E_t^{Q^T}[\frac{S_T}{F(t,T)}] = 1$.Intuitively, following Equation(4.19), the price bounds under forward measure is

$$0 < C(k, \tau) < 1 \tag{4.23}$$

Similarly, I can prove (C3) that $k \to \infty$, the option becomes worthless. From practical point of view, investor would not buy this option because it is impossible to exercise it. C(5) is easy to prove follow the convexity and monotonicity of call option price.

Following Aït-Sahalia and Duarte (2003) and Fengler and Hin (2015), I argue that using the homogeneity assumption, C(1),C(2) and C(4) are sufficient to restrict an arbitrage free call price surface. Therefore in Equation(5.30), I only consider incorporate C(1),C(2) and C(4) in machine learning framework $L_1 - SVM$.

4.4 Summary

In this chapter, I present a new approach of data filter based on three principles: representative, accurate and no arbitrage. Interestingly, as argue by Constantinides et al. (2013), previous studies seem relatively arbitrary because they only consider part of filters. For example, Bakshi et al. (2003) only consider the zero bid price and volatility arbitrage. Jiang and Tian (2005) and Polkovnichenko and

Zhao (2013) restrict the average of bid and ask price. Christoffersen et al. (2013) rule out the zero trading volume. My dataset include S&P 500 option prices January 5, 2000 to April 30, 2016. Since there are no market data for the daily dividend yield, I first use a linear regression to back out the dividend. Although my estimated framework under forward measure is independent from dividend and interest rate, I need them to calculate the forward price and filter dataset. After applying three levels of filter, my dataset ends with 390,320 quotes and 915 firms in sample period.

In addition, I change the measure of the estimated framework. Using Radon-Nikodym derivative, I show that under forward measure, the estimated framework is independent of dividend yield and interest rate. Also, in the last section, I provide the no arbitrage conditions of call option price from the state price density perspective.

Chapter 5

Universal Arbitrage-free Estimation of State Price Density

5.1 Introduction

> "The future has to be based on more a dynamic belief in how markets work and how distributions unfold. Most of risk management technology is based on looking backwards not looking forwards and I do believe that there's huge amounts of information in market prices, in particular in option market prices, about what the forward distribution of risks are, at least as gleaned by the market, and so risk management systems . . . . have to move in the direction of forward information which is contained in derivative contracts and not so much in looking in back."
>
> —Myron Scholes (2016)

Although both practitioners and academics are content with pricing securities based on models that make specific assumptions about the evolution of underlying prices (such as the Black-Scholes-Merton model), they also realize that these models do not always completely conform to the facts of the real world. For example, Rubinstein (1985) documented the phenomenon of the implied volatility smile before the crash of 1987, after which researchers detected pronounced deviations from previous smile shapes[32]. The fact that the real world has richer information than is modeled seems obvious, but only recently has this perspective shifted from an assumed underlying process towards observed market prices and implied distributions.

Thanks to the availability of options data and greatly increased computational power, estimating state price density (SPD) using a data-driven approach has gained attention. Due to its forward-looking nature, this density provides information about market participants' expectations on the evolution of the underlying assets as well as their risk preferences. This is helpful in applications such as managing risk (Aït-Sahalia and Lo, 2000), selecting portfolio (Bali and Murray (2013) and DeMiguel et al. (2013)), studying policy events (Jondeau and Rockinger, 2000), and inferring the empirical stochastic discount factor (Jackwerth, 2000 and Bliss and Panigirtzoglou, 2004).

Broadly, the existing approaches for estimating the SPD can be roughly classified into two strands: parametric and non-parametric (see Jackwerth (1999), Yatchew and Härdle (2006) and Figlewski (2008) for comprehensive reviews). The parametric approach first specifies the SPD as a known distribution (e.g. lognormal distribution) with several unknown parameters then calibrates the unknown parameters by minimizing the discrepancy between the fitted and observed data[33]. Three groups of literature have been developed to add flexibility to the fitting process. First are *the Expansion methods* which introduce correction terms to a known distribution function (e.g. Rosenberg (1998), Jondeau and Rockinger (2001) and Rompolis and Tzavalis (2009)). Second are *Generalized distribution methods* which use more flexible distribution functions with skewness and kurto-

[32]According to Black and Scholes(1973), the implied volatility should be constant. In other word, when plot implied volatility against to strike, the implied volatility surface is flat.

[33]Usually this can be achieved by assuming the underlying dynamics.

sis parameters (e.g. Corrado (2001), Lim et al. (2005) and Fabozzi et al. (2009)). Third are the *Mixture methods* which construct a distribution function as a combination of several simple functions (e.g. Melick and Thomas (1997) and Giacomini et al. (2008)). The non-parametric approach, instead of relying on a specific parametric form, interpolates the SPD between the points and selects the best fit from the possible distributions using set criteria. Specifically, *Kernel regression methods* construct the SPD estimator based on neighboring observations using a kernel function and selected bandwidth (e.g. Aït-Sahalia and Lo (1998)). *Curve fitting methods* fit the observed prices or implied volatility by least squares with flexible functions such as spline, and then derive the SPD via the result of Breeden and Litzenberger (1978). Additionally, there are approaches such as *the Maximum entropy methods* (e.g. Rockinger and Jondeau (2002)) and neural network (Ludwig (2015)), that do not properly fit into either strand and are non-parametric in nature.

Although a considerable number of papers has investigated SPD using various methods, Figlewski (2008) nevertheless argues that estimating SPD is still an open question and none of the techniques is clearly superior. As noted by Cont (1998), each method has drawbacks. The expansion and generalized distribution methods may lead to negative tails in SPD. The Mixture methods exhibit thin tails and Maximum entropy methods gives multimodal SPD estimates because they fail to consider constraints on the smoothness. While the kernel regression methods yield a smooth SPD, the estimator converges slowly and is sensitive to bandwidth choice. The curve fitting methods work well for interpolation but often break down when used for extrapolation. These difficulties are not surprising because estimating SPD poses five challenges. First, unlike theoretical options with continuous strike prices, the options are only traded at discrete strikes. For example, strikes for the S&P 500 Index options are usually spaced \$5 apart. Second, Hentschel (2003) suggests that market option data contain noise from various sources, such as the non-synchronous prices. The index prices are measured fifteen minutes apart from option prices. This makes some noise sensitive methods unattractive. Also, there is an issue with matching market prices with average bid and ask prices. Although most papers[34] use these as true option prices, in practice, it is not obvious how to specify the true price since options are traded within the bid-ask spread. Third, since the SPD lies in $[0, +\infty]$, the range of observable data is insufficient to recovery the information in tails. Fourth, SPD should satisfy no-arbitrage constraints across maturities and strikes. For example, the estimated density should be positive and integrate to one. Finally, the estimation of SPD is afflicted by the "curse of differentiation". This means the quality of the SPD estimator will be much worse than the quality of the option price estimator because the differentiation has an amplifying effect on local irregularities. The small irregularities in observed option prices can easily translate into serious irregularities in the SPD, including negative probabilities.

This chapter relies on a machine learning framework to address four challenges[35]. I first apply a new data filter to eliminate the noise from market data and then propose a more effective approach to estimating the option price via support vector machine (SVM), a method based on statistical learning theory (Vapnik (2000), Schölkopf and Smola (2002)). Rather than attempting to solve the least squares problem, I use a ε insensitive loss function to incorporate bid-ask spread naturally and estimate the SPD by differentiating fitted call option prices

[34]Except Figlewski (2008) considers a weighting function to incorporate the bid-ask spread, Monnier (2013) incorporates bid-ask spread constraints and Glaser and Heider (2012)use Gaussian random variables from bid and ask as input call price.

[35]The third challenge involves extrapolate technique which is not discussed in this thesis, Please see Section 7.2.1 for more detail.

to avoid the “curse of differentiation”.

My approach improves upon other methods in several aspects. First, it is a fully nonparametric and has no strong assumption on the prior distribution compared with a parametric approach. Second, it naturally accommodates the information contained in the bid-ask spread by setting an up-bound predicted error, which allow traders to specify according to their own needs. Third, it is possible to incorporate all arbitrage-free constraints into SVM and give an arbitrage free estimator. Fourth, my approach could easily produce a smooth risk neutral density due to the explicit complexity control and has much better generalization capability compared to other non-parametric approaches. Finally, unlike neural network that need a large amount of input and output data to properly train the network, the SVM has a well-known ability of working well in small sample cases. Finally, this is a universal method. The cubic spline (Fengler (2009)) and tensor product estimator (Fengler and Hin (2015)) are special case of my method using cubic spline and B-spline kernel respectively.

The rest of chapter is organized as follows. Section 2 gives an overview of SPD as well as theoretical no-arbitrage constraints. Section 3 introduces briefly SVM and recasts it in no arbitrage context. An example of application is provided in section 4, where I show how to obtain reliable S&P 500 index options data and apply my approach to recovery a well defined call option prices. Section 5 outlines a framework to compare my approach with three other methods. Section 6 concludes.

5.2 SPD and Option Prices

The SPD is an important link and fundamental block in the economics. It can be obtained from either consumption-based asset pricing models [36](such as Lucas(1978)) or the no-arbitrage asset pricing models. In the consumption-based asset pricing model, the SPD is characterized by marginal rates of substitution (often called stochastic discount factors). In the no-arbitrage asset pricing models, the SPD is refereed as risk-neutral density(RND). Breeden and Litzenberger (1978) show that SPD can be obtained by differentiating the call option pricing function twice with respect to strike price.

5.2.1 SPD overview

Suppose I are given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)$, to which a stock price process S_t is adapted. According to the consumption based asset pricing model , the price of an asset today is equal to the net present value of future payoffs. In these models, the risk free rate r is used to map the future random payoffs to today and the current price can be calculated as the expected value of the stochastic discount factor weighted future payoffs. More formally, the asset price at time t with payoff $Z(S_T)$ is

$$P_t = e^{-r\tau} E_t[M(S_T)Z(S_T)] = e^{-r\tau} \int_0^\infty M(S_T)Z(S_T)f(S_T)dS_T \tag{5.1}$$

Where E_t denotes the expectation under real-world probability $\mathbb{P}$. The stochastic discount factor of Lucas (1978) can be represented as $M(S_T) = \frac{U'(C_{t+1})}{U'(C_t)}$ and

[36]A consumption based asset pricing model focus on demand and supply and is a state dependent function. It usually is characterized by a representation of investor's preference and probability of future outcomes in the model. However, arbitrage-free model do not have a specific form. It considers the arbitrage possibly in the market and widely used in option pricing theory.

reflects the investor's willings of substituting consumption. $Z(S_T)$ is the payoff at time T. S_T is a state variable that drives all the changes in the entire economy. $f(S_T)$ is the payoff probability density under $\mathbb{P}$. To link the real-world probability $\mathbb{P}$ with the state price density, I rewrite the Equation (5.1) as:

$$P_t = e^{-r\tau} E_t^*[Z(S_T)] = e^{-r\tau} \int_0^\infty Z(S_T) p^*(S_T) dS_T \tag{5.2}$$

Where $p^*(S_T)$ is the state price density, which is real-world probability $\mathbb{P}$ discounted by investor's preference. As a consequence, SPD is a product of the real-world probability and investor preferences[37].

The other way to construct the SPD is using no-arbitrage model and option pricing theory. By the fundamental theorem of asset pricing, no-arbitrage is equivalent to the existence of a martingale measure (risk-neutral probability measure Q equivalent to P) under which the discounted stock price process is a martingale. Therefore, the call price of an European option can be written as a discounted expectation of its terminal payoff.

$$C(S_t, K, \tau, r, \delta) = e^{-r\tau} E_t^Q[(S_T - K)^+] = e^{-r\tau} \int_0^\infty \max(S_T - K, 0) q^*(S_T) dS_T \tag{5.3}$$

Where S_t is the underlying asset price at time t, E_t^Q denotes the risk-neutral expectation. $q^*(S_T)$ is the risk neutral probability density of underlying asset. K is the strike price, T is the expiry date and $\tau = T - t$ is time to maturity, δ is the corresponding dividend yield. Compare Equation (5.2) and Equation (5.3), $q^*(S_T)$ and $p^*(S_T)$ are equal. The state price density is also called risk neutral density in option pricing theory.

[37]Indeed, the 'real-world probability' can be recovered from SPD estimator (see Chapter 6 for more information).

With respect to this Equation(5.3), Bredden and Lizenberger(1978) show that $q^*(S_T)$ can be obtained from the continuum of all European call option prices by differentiation of the strike twice. Differencing Equation(5.3) with respect to the strike price K. the risk neutral distribution $F(S_T)$ yields :

$$\frac{\partial C(S_t, K, \tau, r, \delta)}{\partial K} = -e^{-r\tau} \int_K^\infty q^*(S_T) dS_T = -e^{-r\tau}(1 - F(S_T)) \tag{5.4}$$

Further differentiating Equation (5.4) respect to strike price, the risk neutral density(or state price density) can be written as

$$q^*(K) = e^{r\tau} \frac{\partial^2 C(K, T)}{\partial K^2} \tag{5.5}$$

Since the state price density is the second derivatives of European call price function, the major step in estimating state price density is to interpolate a smooth and arbitrage free two dimensional call surface.

5.2.2 Practical Considerations

Although the mathematical relationship between call price function and SPD is clear, the practical implementation of calculating SPD involves three problems. First, it is difficult to find a best fitting parametric SPD estimator since it will be highly sensitive to assumptions on the six variables $C(S_t, K, \tau, r, \delta)$. For example, if the dynamics of the stock price S_t follows the arithmetic Brownian motion[38], then the SPD will be inconsistent with those following a classic geometric Brownian motion. Second, without any restrictions or assumptions on variables, there are too many variables to consider in the nonparametric estimator. From a non-

[38]for example, it is reasonable to assumes the spread options whose underlying spread is positive follows the arithmetic brownian motion.

parametric statistics point of view, high dimensional regression is hardly able to achieve asymptotic consistency in practice. Third, both consumption-based asset pricing models and no-arbitrage asset pricing models assume the no-arbitrage condition, a set of no-arbitrage constraints is needed in estimator.

To overcome the first problem, in this chapter I change the risk neutral numeraire $\mathbb{Q}$ to forward measure $\mathbb{Q}^{\mathbb{T}}$[39]. From the first fundamental theorem of asset pricing, the absence of arbitrage means that there is a numeraire pair, (N, Q_N), where a new probability measure can be introduced by choosing a different N. To be more specific, the Radon-Nikodym derivative that changes $\mathbb{Q}$ to $\mathbb{Q}^{\mathbb{T}}$ can be introduced as:

$$\frac{d\mathbb{Q}^{\mathbb{T}}}{d\mathbb{Q}} = \frac{e^{\int_t^T r(t)dt}}{F(t,T)} \tag{5.6}$$

where $F(t,T)$ is forward measure. This transformation enables the estimating framework in this chapter to consider a zero interest and zero dividend rate case and avoid comparing estimating methods on the ability to deal with input parameters (see Chapter 4 for detail). Further, this chapter applies the dimension reduction method of Aït-Sahalia and Lo (1998). By transforming an option on a stock to a option on future[40], I use the forward price $F(t,T) = S_t e^{(r-\delta)\tau}$ to represent the information of stock price S_t, interest rate r and dividend δ then the call price function can be converted into $C(F_{t,\tau}, K, \tau,)$. Also, by assuming the homogeneity of strike and asset price, the call price function can be reformulated as $C(k,\tau)$, where $k = \frac{K}{F(t,T)}$ is called forward moneyness. In conclusion, the call option price is estimated in the forward-money k and time to maturity τ space and the call price is changed to

[39]see Jarrow(1987) and Geman et al(1995) for change of numerarie method. Also, Gope and Fries (2011)called this step as normalization of call price

[40]According to Aït-Sahalia and Lo (1998), if the future and option have the same maturity, then the European option price on stock is equal to the European option on future.

$$C(k,\tau) = \frac{C(K,\tau)e^{r\tau}}{F(t,T)} \tag{5.7}$$

where $F(t,T)$ is forward price. $C(k,\tau)$ is called the pre-processed call price.

Finally, for ensuring that the estimating framework is independent of arbitrage, next I derive the no-arbitrage constraints for the pre-processed call price. The main results are summarized in Proposition 1 and 2.

Proposition 1. *Let $S_t > 0$. Define the function $C : [0,\infty) \times [0,\infty) \to \mathbb{R}$, such that if $C(K,\tau)$ satisfies following conditions*

(C1) (convexity in strike price K) $C(K,\tau)$is convex function in K for all $\tau \geq 0$

(C2)(monotonicity in time to maturity τ) $C(K,\tau)$is non-decreasing in τ for all $K \geq 0$

(C3) (The call price is limit as strike approach to infinity) $lim_{K\to\infty}C(K,\tau) = 0$ for all τ

(C4) (Price bounds) for all $K \geq 0, \tau \geq 0$

$$max(0, S - K) \leq C(K,\tau) \leq S \tag{5.8}$$

(C5) has expiry value $C(K,0) = max(S - K)$ for all K

Then

there exists a non-negative Markov martingale M_τ such that for all $K, \tau \geq 0$

$$C(K,\tau) = E[(M_T - K)^+ | \mathcal{F}_0 \tag{5.9}$$

that M is a non-negative martingale

Following the conditions of Roper (2010) and assuming that $r = \delta = 0$, I derive the no arbitrage conditions of pre-processed call prices as follows:

Proposition 2. *Under the pre-processed call prices framework, it is evident that (C1), (C2) and (C4) imply*

$$0 < C(k, \tau) < 1 \tag{5.10}$$

$$-1 \leq \frac{\partial C(k, \tau)}{\partial k} \leq 0 \tag{5.11}$$

$$\frac{\partial^2 C(k, \tau)}{\partial k^2} \geq 0 \tag{5.12}$$

see Chapter 4 section 4.3.2.

Existing studies have reached a consensus on the necessary and sufficient conditions to guarantee that option prices are arbitrage free and two papers offer a good insight into how to derive the no-arbitrage conditions of the call price surface. From the option strategy perspective, Carr and Madan (2005) show the idea of static arbitrage, which means there is no arbitrage opportunity on the option price surface. They show that excluding the opportunities of gaining from butterfly spread, calendar spread and other conditions are sufficient to define an arbitrage-free option price surface. From the classical mathematical finance perspective, Roper (2010) helpfully summaries the no-arbitrage condition for a call price surface based on the properties of a martingale. I sketch Theorem 2.1 of Roper (2010) in Proposition 8, then simplify the conditions for a pre-processed call price.

5.3 Support Vector Machine Framework

Suppose the market price data set is $\{(x_1(k, \tau), c_1(k, \tau)), ..., (x_i(k, \tau), c_i(k, \tau)) \subset \mathbb{R}^d \times \mathbb{R}\}$, where i denotes the number of observations and d indexed the dimension of the input space. If I consider the (vectored) call price $C(k, \tau)$, the call price function approximation problem becomes finding a function f such that

$$C(k,\tau) = f(k,\tau) + \varepsilon \tag{5.13}$$

where $f(k,\tau)$ is an estimated function, and ε in an error term. In practice, as long as the estimated call price is in the bid ask range[41] (or trader-specified error tolerance), I can consider the estimated price as precise. Thus, ideally a trader-specified error tolerance is preferable in the estimation framework. This intuition coincides with the key idea of support vector machine.

5.3.1 Support Vector Machine

Theoretically, the input call price $C(k,\tau)$ could be approximated by a linear combination of any continuous function. Put simply, I can use any continuous unknown function to connect the known call price dots.

$$f(k,\tau) = \sum_{i=1}^{\infty} \beta_i \phi_i(k)\phi_i(\tau) + b \tag{5.14}$$

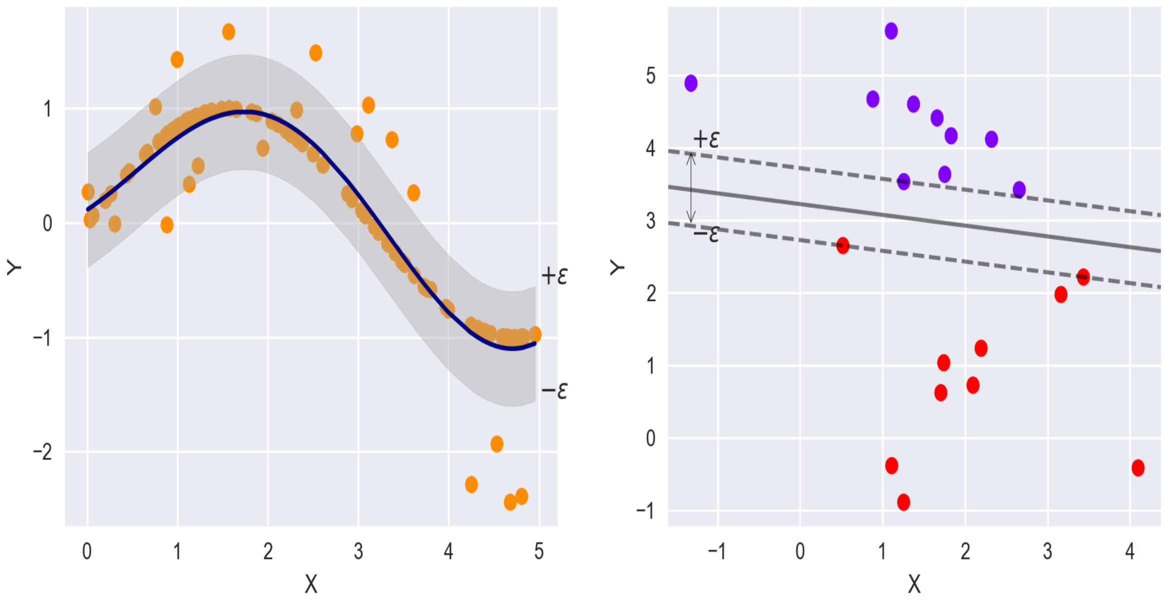
Where $\phi_i(.)$ is basis function such as spline, polynomials, sigmoid function etc. β_i and b are associated coefficients. Motivated by statistical learning theory, the input call price $C(k,\tau)$ can be mapped into a linear feature space by a kernel function. I define the kernel function $K(k,\tau)$ as

$$K(k,\tau) = \phi_i(k)\phi_i(\tau) \tag{5.15}$$

The intuition of SVM is shown in Figure 5.3.1. The left subplot presents the original input space and the right subplot shows the input space after mapping

[41] Insider the [bid price, ask price] range.

Figure 5.3.1: Intuition of Support Vector Machine



Note: The left subplot illustrates the intuition of support vector machine. The right subplot presents the input space after mapping by a kernel function.

by a kernel function. As suggested by the left subplot, the estimated values are allowed have ε discrepancy from the actual call price, or put it another way, I accept the estimated call price inside the gray area[42]. After changing the estimation space via kernel function, as shown on the right in Figure 5.3.1, the estimated function is only influenced by the points near the dashed line. These points are called support vectors. Clearly, a small perturbation of data points away from the dashed line will not affect the slope and shape of the estimated function.

Consequently, I can numerically truncate the Equation (5.14) into

$$f(k,\tau) = \sum_{i \in SV}^{N} \alpha_i K(k,\tau) + b \tag{5.16}$$

[42]The actual call prices are shown by orange points in the left subplot

Where the support vector defines as input pair of (k, τ) which has non-zero associated α and N is the number of support vectors. This remarkable feature turns an estimation problem of finding infinity coefficients β_i to specify a small number of α_i. Moreover, from a practical point of view, I wish the call price function to be explained as simply as possible. Mathematically, this means that $f(k, \tau)$ be as flat as possible and finding the smallest 'l_0 seminorm' of α_i , which has the form

$$min||\alpha_i||_0 \tag{5.17}$$

However, Elad (2010) argues that a better way to address the 'l_0 seminorm' minimization problem is to minimize the l_1 norm because the current algorithm to solve the 'l_0 seminorm' problem is not efficient; thus in my support vector machine framework, my aim turns to finding

$$min||\alpha_i||_1 \tag{5.18}$$

subject to

$$|f(k, \tau) - C(k, \tau)| < \varepsilon \tag{5.19}$$

Where ε can be controlled and bound depending on a trader's need. Furthermore, in practice, a perfect mapping from input data to linear feature space is unobtainable since the true values contain outliers and noise; hence I apply an ε insensitive loss function to penalize the deviation between estimated and true value. As I discussed before, only if a predicted value is outside the bid ask range do I consider it as mis-priced. Formally, the loss function is

$$|\xi|_\varepsilon = \left\{ \begin{array}{cc} 0 & |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & otherwise \end{array} \right. \tag{5.20}$$

The approximation problem is given by

$$min_{(\alpha,b)}||\alpha||_1 + C\sum_{i=1}^{N}|f(k,\tau) - C(k,\tau)| \tag{5.21}$$

subject to

$$-\xi \leq \alpha\mathbf{K} + b - C(k,\tau) \leq \xi \tag{5.22}$$

Where $\boldsymbol{K}$ is a vector of kernel function, the constant $C > 0$ balances the trade-off between the flatness of the estimated model and the amount of deviation allowed. ξ determines the error insensitive zone of the estimated model (the gray tube on the right panel of Figure1). If C is too large, then the objective function (19) is considered to minimize the empirical risk only that means only caring about how well the function approximates the input data. On the other hand, if the ξ is too large, I may get a flat estimated function. In my case, it may result in a noisy call price surface, which entails a multimodal state price density. Thus, the selection of C and ξ is highly important and is usually obtained from gridsearch and cross-validation technique in machine learning theory. In this chapter, I search C from $1e-3$ to $1e3$ and error tolerance ε from $[0, \frac{1}{4}spread]$.

Now I reformulate the estimated problem into matrix form and replacing $C(k,\tau)$ with y, it becomes

$$min_{(\alpha,b,\xi,a)}\mathbf{1}^T\mathbf{a} + C\mathbf{1}^T\xi \tag{5.23}$$

subject to

$$
\begin{aligned}
& -\xi \leq \mathbf{K}\alpha + b\mathbf{1} - y \leq \xi \\
& -\mathbf{a} \leq \alpha \leq \mathbf{a} \\
& 0 \leq \mathbf{1}\varepsilon \leq \xi
\end{aligned}
\tag{5.24}
$$

Where $\boldsymbol{a}$ is a vector of coefficients of $||\alpha_i||_1$.

The linear programing optimization program(5.23) and (5.24) serves as my basic call price SVM estimator. Since the call price function exhibits no arbitrage, further constraints are added in the next section.

5.3.2 No arbitrage Constrained L_1-SVM

As has been seen in section 2.2, the arbitrage free pre-processed call price can be obtained through restrict the first or second order derivatives and output bound of price. In this section, I show that these constraints can be incorporated into the L_1-SVM framework without changing its linear programming nature.

Remark 1. The form of the support vector machine implies that all derivatives of the estimated model are linear.

As noted in Equation (5.16), the estimated call price is linear in the parameters α, if I take first order derivative of the call price respect to the jth component k^j of $k \in \mathbb{R}^d$, the result is

$$
\frac{\partial f(k)}{\partial k^j} = \sum_i \alpha_i \frac{\partial K(\boldsymbol{k}, k_i)}{\partial k^j} = r_1(x)^T \alpha \tag{5.25}
$$

Where $r_1(x) = [\frac{\partial k(k_1)}{\partial k^j} ... \frac{\partial k(k,k_i)}{\partial k^j} ... \frac{\partial k(k,k_N)}{\partial k^j}]^T$. Similarity, if I take the kth order derivative of the call price respect to the jth component k^j of $k \in \mathbb{R}^d$, the result is still linear and only depends on the kernel function and input data

$$f^k(x) = r_k(x)^T \alpha \tag{5.26}$$

Where $r_k(x)$ is the coefficients for the kth order derivative. Let the corresponding value of call price write as vector $\boldsymbol{Y_k}$, The matrix form of monotonicity and convexity constraints $\Gamma_k(Z_k)$ is

$$\Gamma_k(Z_k) = \begin{bmatrix} r_k(x_1)^T & 0 \\ ... & ... \\ r_k(x_p)^T & 0 \\ ... & ... \\ r_k(x_{|Z_k|})^T & 0 \end{bmatrix} \tag{5.27}$$

$$Y_k(Z_k) = \begin{bmatrix} y_1^{(k)} \\ ... \\ y_p^{(k)} \\ ... \\ y_{|Z_k|}^{(k)} \end{bmatrix} \tag{5.28}$$

Where $Z_k = \{k_{1,}k_{2,...,}k_{|Z_k|}\}$ contains interesting points regarding derivative constraints. The choice of Z_k may be vary and depend on a trader's judgement. In my case, I argue that 0 should be included as $c(0,\tau) = \frac{e^{r_\tau}}{F^\tau}$. For the spline kernel, Z_k is called knot[43]. Furthermore, consider the price bounds constraints as a restriction on the zero order derivative of the call price, I establish the general

[43]Please referred Fengler and Hin (2015)for choosing ideal knots using Akaike Information Criterion(AIC).

no arbitrage constrained SVM

$$min_{(\alpha,b,\xi,a)}\mathbf{1}^T\mathbf{a}+C\mathbf{1}^T\xi+\sum_{k=1}^{n_{sc}}\lambda 1^T z_k \tag{5.29}$$

subject to

$$\begin{aligned}
&-\xi \leq \mathbf{K}\alpha + b\mathbf{1} - y \leq \xi\\
&-\mathbf{a} \leq \alpha \leq \mathbf{a}\\
&0 \leq \mathbf{1}\varepsilon \leq \xi\\
&\Gamma_k(Z_k)\theta - Y_k(Z_k) \leq z_k
\end{aligned} \tag{5.30}$$

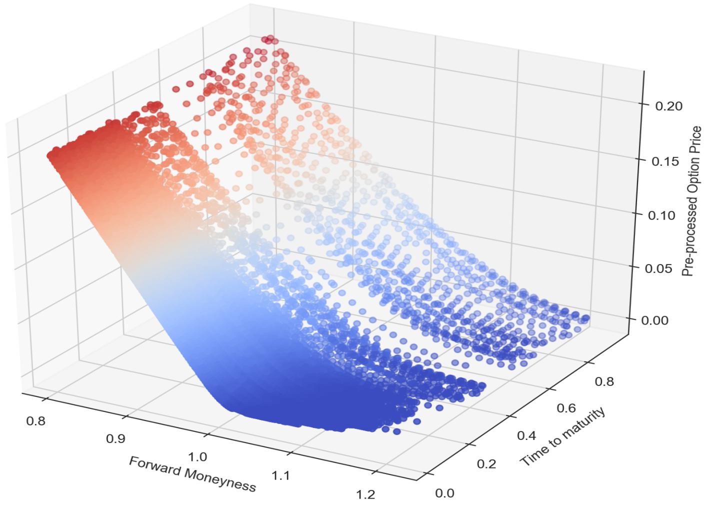
5.4 Empirical Analysis

5.4.1 Data Description

To illustrate the effectiveness of my method, I use daily prices for S&P 500 index options from OptionMetrics. The S&P 500 index is taken as a good indicator for the U.S. market portfolio, and the corresponding options are therefore expected to infer the consensus between market participants. The dataset includes closing index price and interest rate. As regards the option price, I collect the bid and ask price, trading volume and strike price. The dataset run from January 5, 2000 to April 30, 2016, which yields 4,025 trading dates and 412 expiration dates. The average daily trading volume of each contract is 1321.35. My dataset contains total 8,261,170 quotes with time to maturity varying from 7 days to 365 days. I use the average of the bid and ask price as 'true' call price, but I set ε in Equation (5.30) as a quarter of bid and ask spread to consider the information within the bid and ask range.

Unlike a parametric method, my non-parametric L_1-SVM estimator is data-driven and requires arbitrage-free input data. Therefore, my dataset poses three

Figure 5.4.1: Scatter Plot of Pre-processed Call Option Price



Note: This figure plots the pre-processed call price from 01/01/2016 to 31/04/2016. The x-axis is forward moneyness and y-axis is time to maturity.

challenges. First, the option prices are imprecise because tick sizes, bid-ask spread, and non-synchronicity of index and option prices constitute a source of error (Hentschel (2003)). The true trade price is not always centered between the bid and ask prices. Second, there are no observable data for the daily dividend yield. Third, in the money (ITM) options are less liquid than out the money (OTM) options, with a potential impact of differential liquidity on prices.

To ensure that I have reliable option quotes and solve the above challenges, I apply three levels of the filter in Chapter 4 to eliminate the influence of liquidity and errors. I first remove identical observations from the OptionMetrics database then select liquidity data using certain criteria. Finally, I delete outliers based on the value of implied volatility and implied interest rate. To solve the second challenge and increase data quality, I first linearly interpolate the interest rate and use the put-call parity relationship to derive the implied dividend (see Chapter 4

Table 5.4.1: Descriptive Statistics of S&P 500 OTM Options Data

Panel A: Descriptive Statistics

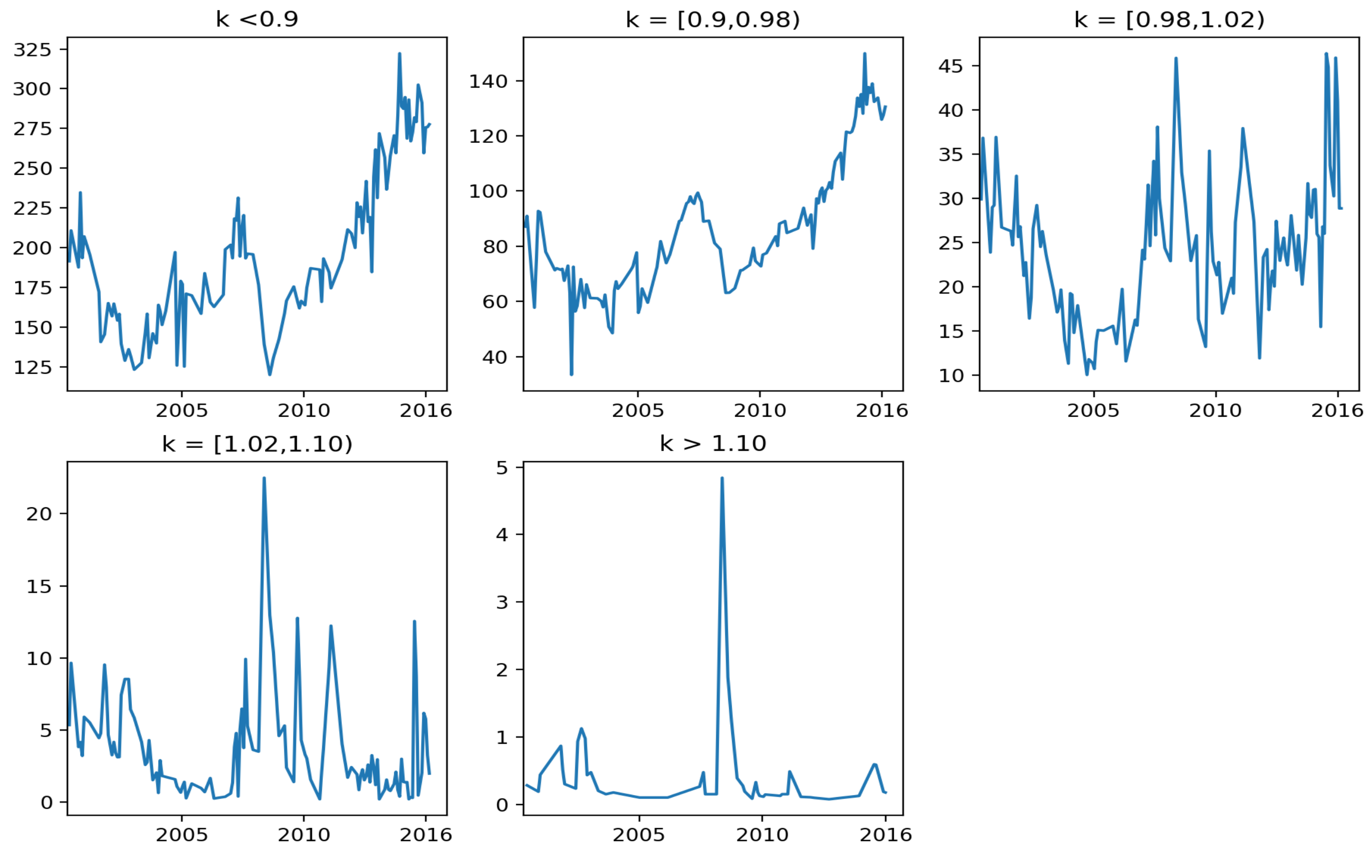
Variables	Mean	Std. Dev.	Min	Percentiles 5%	10%	50%	90%	95%	Max
Call Price ($)	13.31	15.13	0.08	0.28	0.50	8.15	33.90	43.75	146.3
Put Price ($)	12.77	14.31	0.08	0.50	0.93	7.60	32.20	42.05	151.50
Implied Volatility σ (%)	20.62	8.39	6.46	10.30	11.46	19.19	31.10	35.69	87.11
Strike price(K)	1537.75	377.62	550	930	1030	1540	2030	2110	2500
Days(τ)	53.82	41.47	7	14	18	44	95	127	365
Index Price (S)	1601.16	389.84	676.53	944.89	1091.60	1606.28	2079.51	2098.48	2130.82
Trading Volume (V)	1321.35	3690.84	1	2	4	110	3581	6698	157542

Panel B. Options by Forward Moneyness-Maturity

Forward Moneyness K/F	DTM <0.90	ITM [0.90, 0.98)	ATM [0.98, 1.02)	OTM [1.02, 1.10)	DOTM >1.10
	Average Call Prices ($)				
Short-term $\tau \in (7, 60]$	237.04	107.40	30.30	6.60	2.05
Medium-term $\tau \in (60, 180]$	243.74	120.17	50.02	15.35	4.60
Long-term $\tau \in (180, 365]$	249.55	144.49	85.76	42.38	12.96
	Average Implied Volatility σ(%)				
Short-term $\tau \in (7, 60]$	29.96	20.92	16.05	15.54	22.36
Medium-term $\tau \in (60, 180]$	25.26	19.93	16.75	14.51	16.82
Long-term $\tau \in (180, 365]$	22.59	19.46	17.51	15.70	14.39

Note: Table 5.4.1 summaries descriptive statistics of my S&P 500 option dataset (total 390,320 observations). The sample period is from January 5, 2000 to April 30, 2016. The call price is calculated as average of bid and ask price. The OTM put prices are translated into ITM call prices by put and call parity. Std.Dev denotes the standard deviation from call option price.

Figure 5.4.2: Monthly Average of 1-month Call Option Prices over Different Forward-moneyness



Note: This figure plots the time series of call price between 2000 and 2016 across different forward moneyness groups. k is forward-moneyness.

for details). Over the sample period, the average risk-free rate and implied dividends were 1.17% and 2.53%, respectively. After calculating the implied continuous dividend, I compute the Black-Scholes implied volatility using the Bisection approach. Finally, I obtain 390,320 observations[44]. The resulting processed call prices are free of assumptions on interest rate and dividend yield.

Figure 5.4.1 shows the pre-processed call price of final dataset from 01/01/2016 to 31/04/2016. Clearly, my final dataset yields a well-defined call option price surface. Panel A in Table 5.4.1 presents descriptive statistics for my sample of S&P 500 daily prices and Panel B reports the statistics of call prices and implied volatility by time to maturity and forward-moneyness groups. As expected, implied volatility shows a clear smile pattern and call price decreases as forward moneyness increases. Figure 5.4.2 compares the behavior of the monthly average of 1 month S&P 500 option prices over different froward moneyness. Consistent with Barletta et al. (2017), not surprisingly, the ITM options exhibit non-stationary behavior, whereas the OTM options are stationary. Barletta et al. (2017) further explain this as the price of ITM option mainly associated with the underlying index since holding a price of the call option is equal to holding an underlying index strike price approaching zero. In other words, the behavior of the ITM option price is similar to the S&P 500 trend to some extent.

5.4.2 Performance Measures

When comparing the proposed L_1-SVM method with other nonparametric methods (Fengler (2009) and Fengler and Hin (2015)), I assess with regard to three aspects: accuracy of estimation, running speed and smoothness of call price surface with respect to forward-moneyness and time to maturity. Although the mean

[44] This is a reasonable number compares with Chiang etc(2016) who get total 404,822 observations using the same filters on S&P 500 options between 1996 and 2011.

squared error (MSE) is a general accuracy indicator, in this thesis, I use relative distance to measure the goodness of fit. This is because the total estimated number of my L_1-SVM method is different from (Fengler (2009) and Fengler and Hin (2015). As proposed in Section 5.4, I use a 5 fold cross-validation process to determine the optimal parameter and calculate the estimated error, and so the final estimated number equals the number of its training subset. Put simply, if I choose 5 fold cross-validation, the number of total estimated option price only accounts 20% of the whole data size. Considering this size effect, I use a relative error to measure the estimated goodness.

- **Accuracy**

I define the relative error as relative distance

$$Relative \quad Distance = \sqrt{\sum_{n=1}^{N} (\frac{C(k,\tau) - \hat{C}(k,\tau)}{C(k,\tau)})^2} \tag{5.31}$$

where $C(k, \tau)$ is pre-processed call price from market and $\hat{C}(k, \tau)$ estimated by different methods. N is the number of total estimated option price.

- **Smoothness**

I use the absolute value of second order derivative for both dimensions to measure the smoothness of the surface. I use a numerical differentiation approach, applying the central finite difference approximation.

$$\frac{\partial^2 \hat{C}(k,\tau)}{\partial k^2} \approx \frac{C(k_{m+1},\tau) - 2C(k_m,\tau) + C(k_{m-1},\tau)}{(k_m - k_{m-1})(k_{m+1} - k_m)} \tag{5.32}$$

$$\frac{\partial^2 \hat{C}(k,\tau)}{\partial \tau^2} \approx \frac{C(k,\tau_{m+1}) - 2C(k,\tau_m) + C(k,\tau_{m-1})}{(\tau_m - \tau_{m-1})(\tau_{m+1} - \tau_m)} \tag{5.33}$$

Where $1 < m < N$. The smaller the second order derivatives, the smoother the call option price surface is.

5.4.3 Empirical Results

Using the pre-processed S&P 500 option prices, I first compare the proposed L_1-SVM method using a radial basis function (RBF) and a spline kernel[45]. The algorithm for the full estimating process is summarized in Algorithm 1. As noted in Section 3.1, the algorithm's performance is determined by the parameter choice for the kernel function. I apply a grid search with 5-fold cross validation[46]to find the optimal parameters that minimize Equation(5.28). Table 5.4.3 reports the mean relative distance of specific kernel functions. I test the univariate and bivariate approximation using RBF and cubic spline kernel in 3 sub-periods (the period before the financial crisis, financial crisis period and the period after the financial crisis). In each case, I report the relative distance of L_1-SVM. The cubic spline and tensor product B-spline function is introduced in Section 3.1.2.

I expect intuitively that the relative distance for the financial crisis period is higher because the financial market is highly volatile in this period. Therefore, the L_1-SVM needs additional effort to fit these observations. However, as shown in Table 5.4.3, when comparing the relative distance of the L_1-SVM across different periods, the relative distance turns out to be decreased in the financial crisis period for both kernels in univariate and bivariate cases. This suggests that my L_1-SVM method is not affected by volatility in the market. The smallest relative distance is obtained by using the univariate cubic spline kernel. Also, consistent with Fengler and Hin (2015), I find that the tensor product kernel (or

[45] The kernel function is reported in Section 3.1.2 and Appendix.

[46] I use the Python library GridSearchCV and Cvxopt to program the code and the laptop conducts the codes is an Intel Core i7 and 2.9 GHz.

Algorithm 1: Approximation call price surface use support vector machine

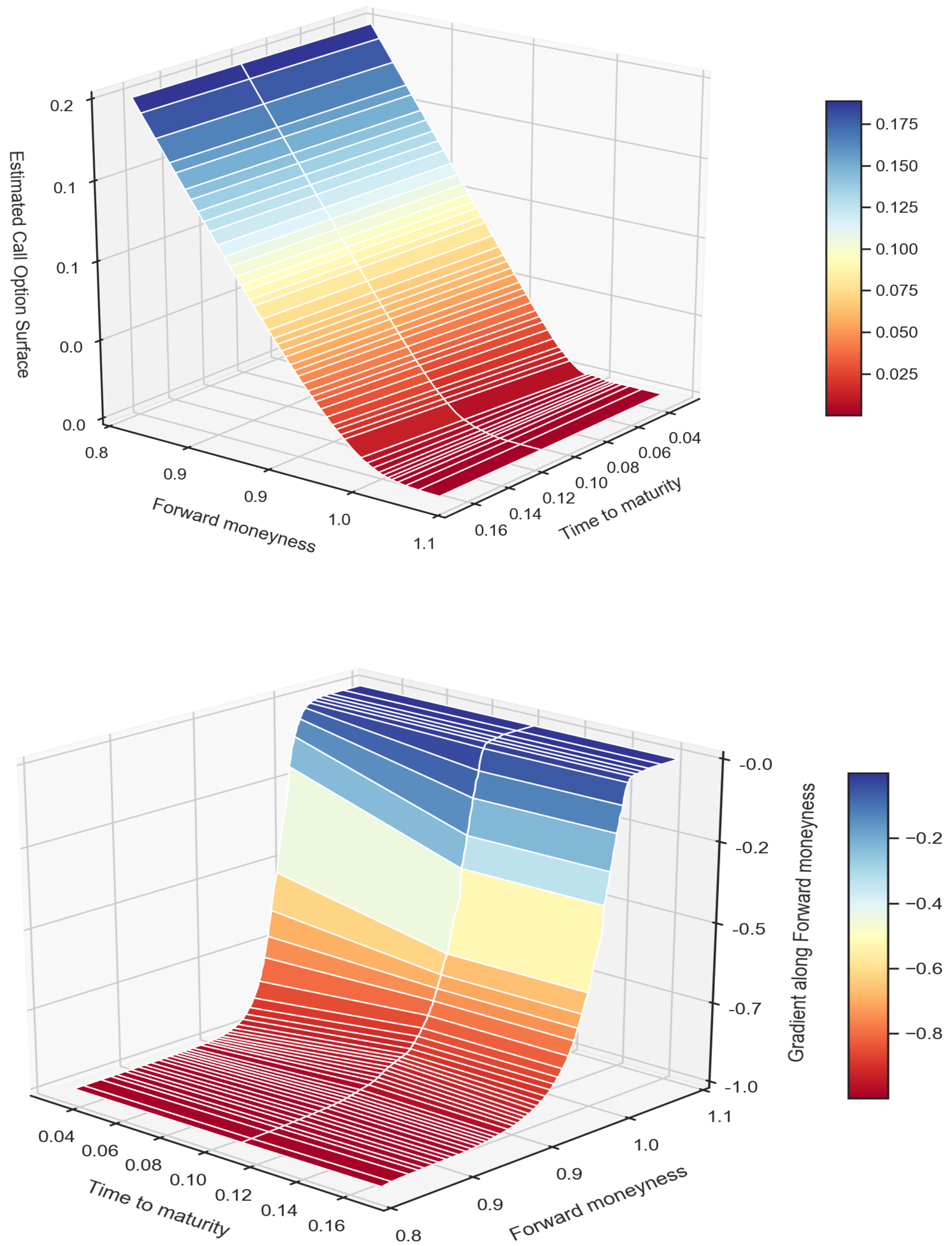
(1) Initial ;
Input: Observed forward moneyness κ_i, i = 1,...N
Observed maturity τ_i, i = 1,...N
Observed European call price $c_i(\kappa_i, \tau_i)$
a. Applying the three levels data filters in Chapter 4
b. Transform option prices $c_i(K_i, \tau_i)$ to pre-processed call prices(under forward measure) $c_i(\kappa_i, \tau_i)$
c. Estimate and compare call price under forward measure
d. Transform the estimated call price under risk neutral measure Q
e. Estimate the state price density
Output: Estimated call price $c_i(\hat{\kappa}, \hat{\tau})$

Estimate the call price under forward measure $c_i(\hat{\kappa}, \hat{\tau})$
(1) Randomly split data D into 5 "folds" of equal size: $D_1, D_2....D_5$
(2) For i = 1,....5,
$min_(\alpha, b, \xi, a) \quad 1^T\alpha + C1^T\xi + \sum \lambda 1^T z_k$
subject to Equation(5.30)
fit above model use defined parameters set of C, ε and γ (for different kernel parameters)
(3) Return optimal parameters of C, ε and γ, which lead the minimal test error
(4) Use optimal parameters to fit the call option price function

called surface estimator in Fengler and Hin (2015)) seems to underperform the univariate kernel (a slice by slice approach). In my result, the relative distances for the bivariate kernel approximation are 2.732 and 2.359 for RBF and cubic spline in the period after the financial crisis respectively, which is significantly higher than the corresponding univariate case. Fengler and Hin (2015) argue that without considering calendar-spread arbitrage (Equation(2.28)-(3.30)), the surface estimator does not improve the fitting quality. In other words, without using calendar-spread constraints, using a surface estimator only increases the fitting difficulty since it uses a base surface to fit the call option price surface.

Figure 5.4.3 displays an example of the estimated arbitrage free call option

Figure 5.4.3: Estimated Call Option Price and Its First-order Derivative



Note: Figure 5.4.3 plots the estimated call price and its first-order derivative on 02/07/2013. The index price is 1614.08. The top panel shows the estimated call price. The bottom panel displays the first-order derivative of estimated call price respect to forward-moneyness.

Table 5.4.3: Empirical Results with RBF and Spline Kernel

Period	Univariate		Bivariate	
	RBF	Cubic Spline	RBF	Cubic Spline
2000-2007	4.664	2.894	5.114	3.086
2008-2009	3.934	1.737	3.945	2.645
2010-2016	2.805	1.032	2.732	2.359

Note: This table compares the performance of L_1-SVM of radial basis function(RBF) and cubic spline kernel. For each kernel, I compare the relative distance of the univariate and bivariate case. Univariate refers as using a kernel function of forward-moneyness to fit each maturity slice. Bivariate refers as using a tensor product kernel function of forward-moneyness and maturity to the pre-processed call price surface.

price and its first order derivative. Consistent with Equation(5.10) and (5.11), the call option price under the forward measure is greater than 0 and less than 1. Its first-order derivative monotonically increases with forward-moneyness. If I reverse the change of measure (Equation(5.6), this first-order derivative under risk neutral density is called delta, which measures the sensitivity of option price to change in the underlying value.

5.5 Comparison of nonparametric methods

In the previous sections, I test the proposed method using different kernels and show the cubic spline kernel yields the smallest relative distance. To asses my machine learning based framework with another nonparametric methods, I first summarize the differences between these nonparametric methods into two aspects:

- Assumptions of input variables: as shown in Section 5.2.2, the call price function can be expressed as $C(S_t, K, \tau, r, \delta)$, in which the variables S_t,K and τ can be easily obtained from the market while r and δ are difficult to calibrate. Without using the forward measure as in this thesis, previous studies have different assumptions regards these two variables (such as Glaser and Heider (2012) assume they are both constant).

- Choice of approximation method includes two important parts: choice of kernel function and kernel dimension.
  - Various kernel functions have been applied to approximate the call option price surface in previous studies such as low order polynomial (Kundu et al. (2016)), radial basis function (Lai (2011)) and cubic spline kernel (Fengler (2009)). In this chapter, since the smallest relative distance is obtained by using a cubic spline kernel, I compare my method with Fengler (2009).
  - In terms of kernel dimension, Fengler and Hin (2015) is the only study that uses the bivariate kernel approximation. In contrast with this thesis, they incorporate the no-arbitrage constraints in the control net of the tensor product B-spline and solve a quadratic programming to fit the call option price surface.

To compare my L_1-SVM method with Fengler (2009) and Fengler and Hin (2015), I first summarize the difference between two methods in Table 5.5.1. Since my estimated framework is independent of interest rate and dividend yield, the performances of these three models are mainly determined by their optimization procedure. I briefly review the key optimization function of Fengler (2009) and Fengler and Hin (2015) as follows.

- Maturity Slice by Slice (Fengler (2009))

Dividing the call option price surface into several maturity slices, Fengler (2009) proposes a method that estimates the call option price using a cubic spline function. To make the call option price surface as smooth as possible, Fengler (2009) first develops a smooth technique in implied volatility space. After obtaining the smoothed data, he transforms them back to the option price space. Using the no-arbitrage conditions on the option price, he simplifies the estimation problem as follow:

For each maturity

$$min_{g_\tau} \sum_{m=1}^{M} w_m[c(k_m,\tau) - g(k_m)]^2 + \lambda \int_{k_1}^{k_m} (g''_\tau(k))^2 dx \tag{5.34}$$

Where $g(k_m)$ is spline series, $c(k_m,\tau)$ is the observed call price. $g''(k)$ is a regularization term introduced by Green and Silverman (1994). The optimization problem Equation(5.31) is solved with respect to the no-arbitrage conditions. I present the matrix form of final estimation problem in Algorithm 2[47].

- Two dimensional Tensor Product Kernel (Fengler and Hin (2015))

Fengler and Hin (2015) extends the earlier method using a two univariate spline kernel. This method represents the call option price surface as a linear combination of tensor product B-spline surface. By minimizing a penalized least squares , the estimated problem becomes

$$min_\theta \frac{1}{N} \sum_{i=1}^{N} (c(\kappa_i,\tau_i) - s(\kappa_i,\tau_i))^2 + \lambda_N |\theta|^2 \tag{5.35}$$

where $s(k_i,\tau_i)$ is tensor product spline, $c(k_i,\tau_i)$ is the observed call price. θ is vector of the tensor product B-spline coefficients. Without directly considering no-arbitrage in the quadratic programming framework, Fengler and Hin (2015) establish no-arbitrage conditions on an artificial kernel surface (which is called

[47]Please refer to Fengler (2009) for more detail

the control net of the B-spline). This approach involves a complex knot search and relocate process, which is highly time consuming. I present the matrix form of this method in Algorithm 3[48].

5.5.1 Comparison Results and State Price Density

Table 5.5.2 reports the comparison results for these three methods. Not surprisingly, my L_1-SVM method shows similar result to those of Fengler (2009) because they both apply a univariate cubic spline kernel. This similar result strongly indicates that my L_1-SVM method is a universal approach. Previous studies that use different kernel and optimization procedures can be replicated by my method. As shown in the table, compared to Fengler (2009), although my L_1-SVM method is slow, it shows a somewhat better accuracy. I argue that this can be attributed to my comprehensive data filter approach, simple linear programming framework, incorporating the bid-ask spread information or all of them.

- Analysis of computation time

In Figure 5.5.1, I show the evolution of computation time during the estimated period. It is evident that Fengler and Hin (2015)'s running time far exceeds the others. As shown in Algorithm 3, this is because it searches simultaneous optimizations for forward-moneyness and maturity. Unlike the other two methods, which are completely unfazed by the financial crisis period, I observe a significant peak of Fengler and Hin (2015) around the financial crisis period, which implies that Fengler and Hin (2015) may need additional effort to calculate more volatile data. In the recent period, my method and Fengler (2009) both highlight two notable peaks, and, as both methods are based on the maturity slice approach, I argue that these may be caused by the market microstructure; for example,

[48]Please refer to Fengler and Hin (2015) for more detail

possible price manipulation through controlling order size, or my data filter does not eliminate all possible arbitrary data. I suggest this for future research.

- Analysis of relative distance

For the relative distance, as shown in Tale 5.1.2, the mean of relative distance for Fengler and Hin (2015), Fengler (2009) and my L_1-SVM method is 18.734, 2.368 and 2.363 respectively. My L_1-SVM method produces the smallest relative distance and Fengler and Hin (2015) displays the worst performance with the mean of relative distance almost 8 times higher. Figure 5.5.2 shows the changes of relative distance over time. My L_1-SVM method shows the similar pattern with Fengler (2009) but with slightly lower value.

- Analysis of smoothness

When comparing the smoothness of call price surface in the forward-moneyness direction, my L_1-SVM method stands out by having a stable result. As illustrated in Figure 5.5.3, without considering the data around 2016, my L_1-SVM method exhibits stable behavior with a lower absolute second-order derivative with respect to forward-moneyness. Another interesting comparison among the three methods is to compare the surface smoothness in the time to maturity direction: it can be seen in Table 5.1.2 that there are no significant differences in their ability to interpolate across time. Except the max value and standard deviation, my method shows the same result as Fengler (2009). This finding provides empirical evidence for Fengler and Hin (2015)'s simulation result, which shows the calendar

Algorithm 2: Fengler(2009): Approximation of the call price surface with cubic spline

(1) Calculate pre-estimated call price surface;

Input: Observed forward moneyness level κ_i, i = 1,...N
Observed maturity level τ_i, i = 1,...N
Observed Black Schole implied volatility $\sigma_i(\kappa_i, \tau_i)$

a. Define a regular grid as $\hat{\kappa} \times \hat{\tau} \subseteq [\kappa_i, \tau_i]$ with 100 grid points

b. Get pre-estimted implied volaitlity
$\sigma_i(\hat{\kappa}, \hat{\tau}) = \textbf{Interpolate}(\sigma_i(\kappa_i, \tau_i))$

Output: Estimated call price $c_i(\hat{\kappa}, \hat{\tau})$)

(2) Fitted the call price surface slice by slice(from last to the frist maturity), and solve the quadratic program. For t_m, solve

$$Q = \begin{bmatrix} \frac{1}{h_1} & 0 & 0 & 0 & 0 \\ (-\frac{1}{h_1} - \frac{1}{h_2}) & \frac{1}{h_2} & 0 & 0 & 0 \\ \frac{1}{h_2} & (-\frac{1}{h_2} - \frac{1}{h_3}) & \dots & \dots & 0 \\ 0 & \frac{1}{h_3} & \dots & \dots & 0 \\ \vdots & 0 & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & \frac{1}{h_{M-2}} \\ 0 & 0 & 0 & 0 & (-\frac{1}{h_{M-2}} - \frac{1}{h_{M-1}}) \\ 0 & 0 & 0 & 0 & \frac{1}{h_{M-1}} \end{bmatrix}$$

$$R = \begin{bmatrix} \frac{1}{3}(h_1 + h_2) & \frac{1}{6}h_2 & 0 & 0 & 0 \\ \frac{1}{6}h_2 & \frac{1}{3}(h_2 + h_3) & \frac{1}{6}h_3 & 0 & 0 \\ 0 & \dots & \dots\dots & 0 & \\ 0 & 0 & \frac{1}{6}h_{M-3} & \frac{1}{3}(h_{M-3} + h_{M-2}) & \frac{1}{6}h_{M-2} \\ 0 & 0 & 0 & \frac{1}{6}h_{M-2} & \frac{1}{3}(h_{M-2} + h_{M-1}) \end{bmatrix}$$

Define $A = (Q, -R^T)$ and set $W_M = diag(w_1, w_2...w_M)$,where
$h_m = \hat{\kappa_{m+1}} - \hat{\kappa_m}$

$$B = \begin{pmatrix} W_M & 0 \\ 0 & \gamma \end{pmatrix}$$

For each time to maturity:
Set: $x = (g^T, \gamma^T)^T$ and $y = (w_1 c(\hat{\kappa_1}, \hat{\tau_1}), ..., w_M c(\hat{\kappa_M}, \hat{\tau_l}), 0, ...0)^T$
$min_x \quad -y^T + \frac{1}{2}X^T BX$
Subject to $A^T X = 0$
$\hat{c}(\hat{\kappa}, \hat{\tau_l}) = g^*_{\tau_l}$

Algorithm 3: Fengler and Hin(2013): Approximation of the call price surface with tensor product B spline

(1) Initial ;

Input: Observed forward moneyness κ_i, i = 1,...N

Observed maturity τ_i, i = 1,...N

Smoothing parameter λ

Observed European forward call price $c_i(\kappa_i, \tau_i)$

Output: Estimated call price $c_i(\hat{\kappa}, \hat{\tau})$

(2) Fitted the call price surface using tensor product of spline kernel

a. Define the tensor product kernel as

$$B = \begin{bmatrix} (B_{0,p_1}(\kappa_1))...(B_{q_1,p_1}(\kappa_1)) \bigotimes (B_{0,p_1}(\tau_1))...(B_{q_1,p_1}(\tau_1)) \\ (B_{0,p_1}(\kappa_N))...(B_{q_1,p_1}(\kappa_N)) \bigotimes (B_{0,p_1}(\tau_N))...(B_{q_1,p_1}(\tau_N)) \end{bmatrix}$$

Where p_1 and p_2 are the order of b spline kernel

b. Define a regular grid as $\hat{\kappa} \times \hat{\tau} \subseteq [\kappa_i, \tau_i]$ with 100 grid points

c. Search and Relocate the knots in grid that have minimum Akaike Information Criterion

The search boundary is v and the new knot as ϵ^{add}, the orignal knots is ϵ^0 and the AIC

value of original knots is AIC_0

While the knots in grid do:

$\epsilon^{add} = argmin(AIC(\epsilon^0 \bigcup v)$

Until

$AIC(\epsilon^{add} \bigcup \epsilon^0) > AIC_0$

(3)Set optimization problem as $D = B^T B + \lambda I$

$$B^T = \begin{pmatrix} c_1(\kappa_1, \tau_1) \\ c_1(\kappa_N, \tau_N) \end{pmatrix}$$

Solve the quadratic program

$min(0.5\theta^T D\theta - \theta^T d)$

subject to constraints on coefficients of b spline kernel(reference papr for more detial)

Finally

$\hat{c}(\hat{\kappa}, \hat{\tau}_l) = \sum_{j_1=0}^{q1} \sum_{j_2=0}^{q2} \theta_{j_1,j_2} B_{j_1,p_1}(\kappa) B_{j_2,p_2}(\tau)$

Table 5.5.1: Summary of Compared Models

Model	Interest rate	Dividend	Kernel	Scope
Fengler(2009)	Deterministic	Deterministic	Cubic Spline	Maturity Slice (Univariate)
Fengler and Hin (2013)	Deterministic	Deterministic	Tensor product B-spline	Global (Bivariate)

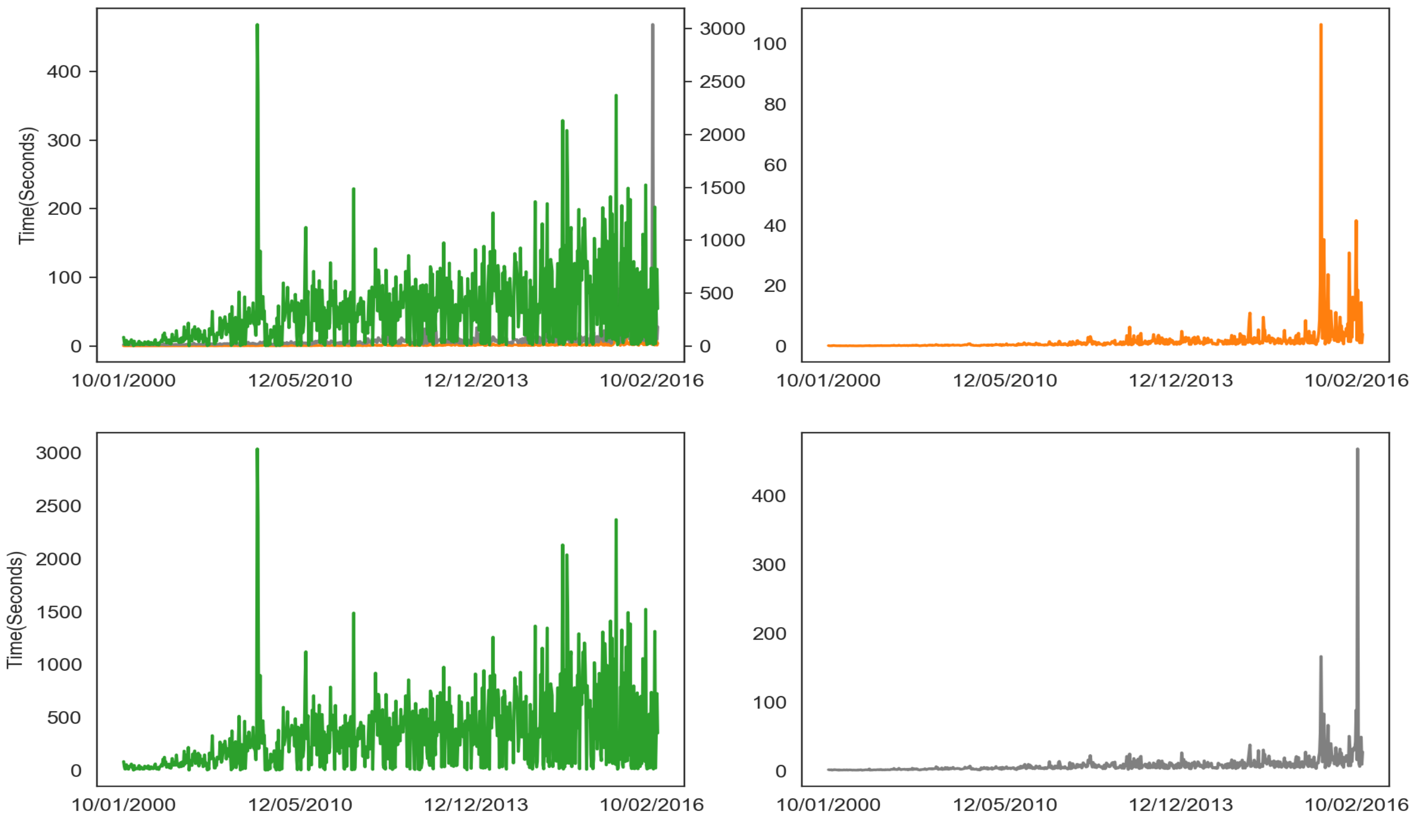
Note: this table compares two nonparametric estimation methods for assumption of interest rate and dividend, choice of kernel in my estimation framework.

Table 5.5.2: Empirical Results for Estimating Call Option Price

		Fengler (2009)	L_1-SVM	Fengler (2015)
	Min	0.093	1.259	4.579
Time	Mean	1.708	8.771	331.934
	Max	106.481	468.83	3037.909
	Std	5.043	19.680	305.849
	Min	0.314	0.312	1.164
Relative Distance	Mean	2.368	2.363	18.734
	Max	10.132	10.135	367.530
	Std	1.933	1.937	27.119
	Min	0.060	0.004	0.082
Smoothness(Moneyness)	Mean	35.250	33.369	43.633
	Max	4406.122	2854.166	1197.856
	Std	171.411	184.698	97.831
	Min	0.000	0.000	0.000
Smoothness(TTM)	Mean	0.006	0.006	0.010
	Max	0.116	0.117	0.231
	Std	0.011	0.016	0.020

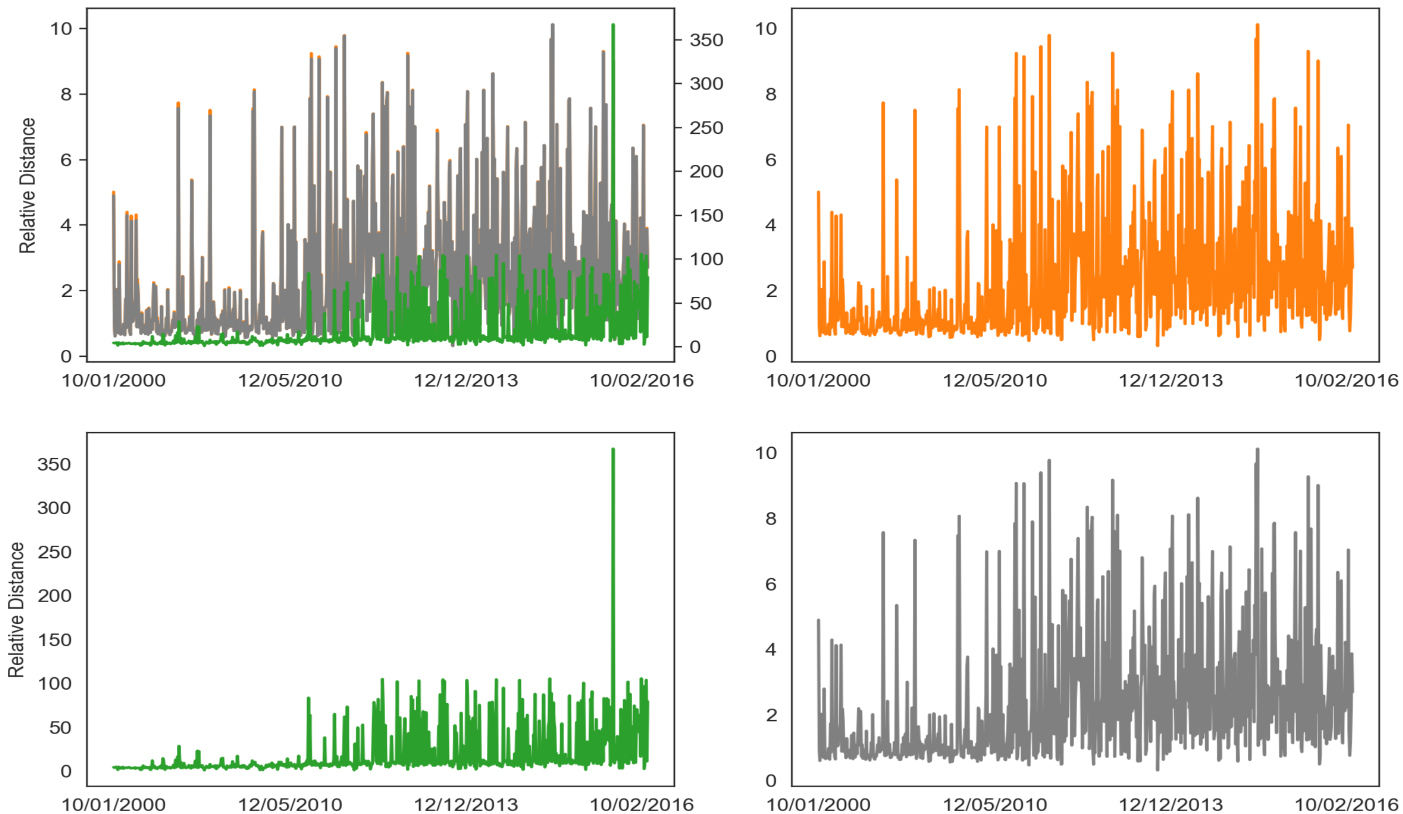
Note: this table reports the performance of three compared methods. Time refers as the computation time. Relative distance and smoothness (forward-moneyness and maturity direction) is defined in Section 5.5.1. Std represents standard deviation. I use Python's Cvxopt library to solve the optimization and the laptop conducts the codes is an Intel Core i7 and 2.9 GHz.

Figure 5.5.1: Running Time Comparison



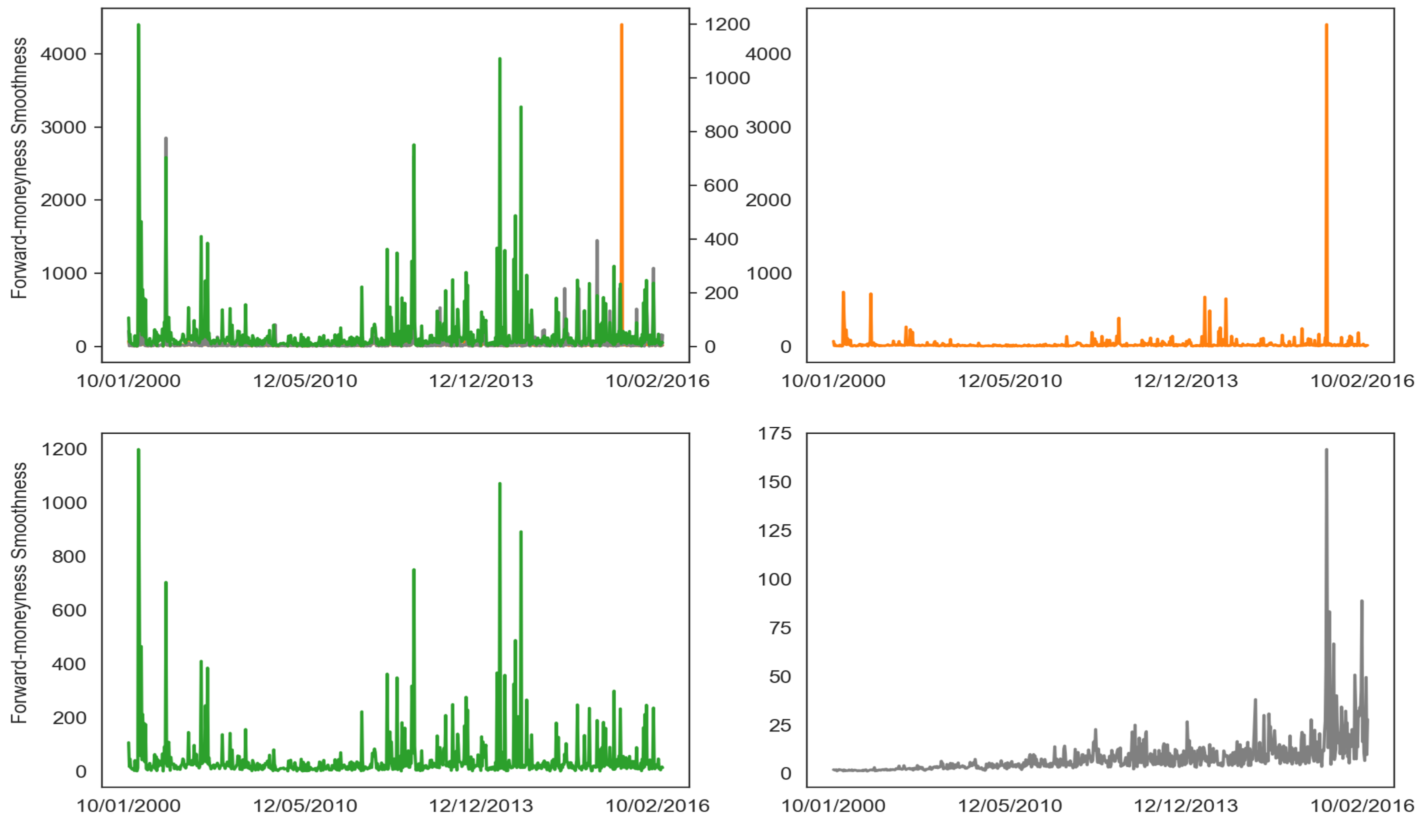
Note: this figure displays the computation time of three methods. The top left panel plots all results. The green line represents Fengler and Hin (2015). The orange line represents Fengler (2009) and the gray line represents the result of L_1-SVM.

Figure 5.5.2: Relative Distance Comparison



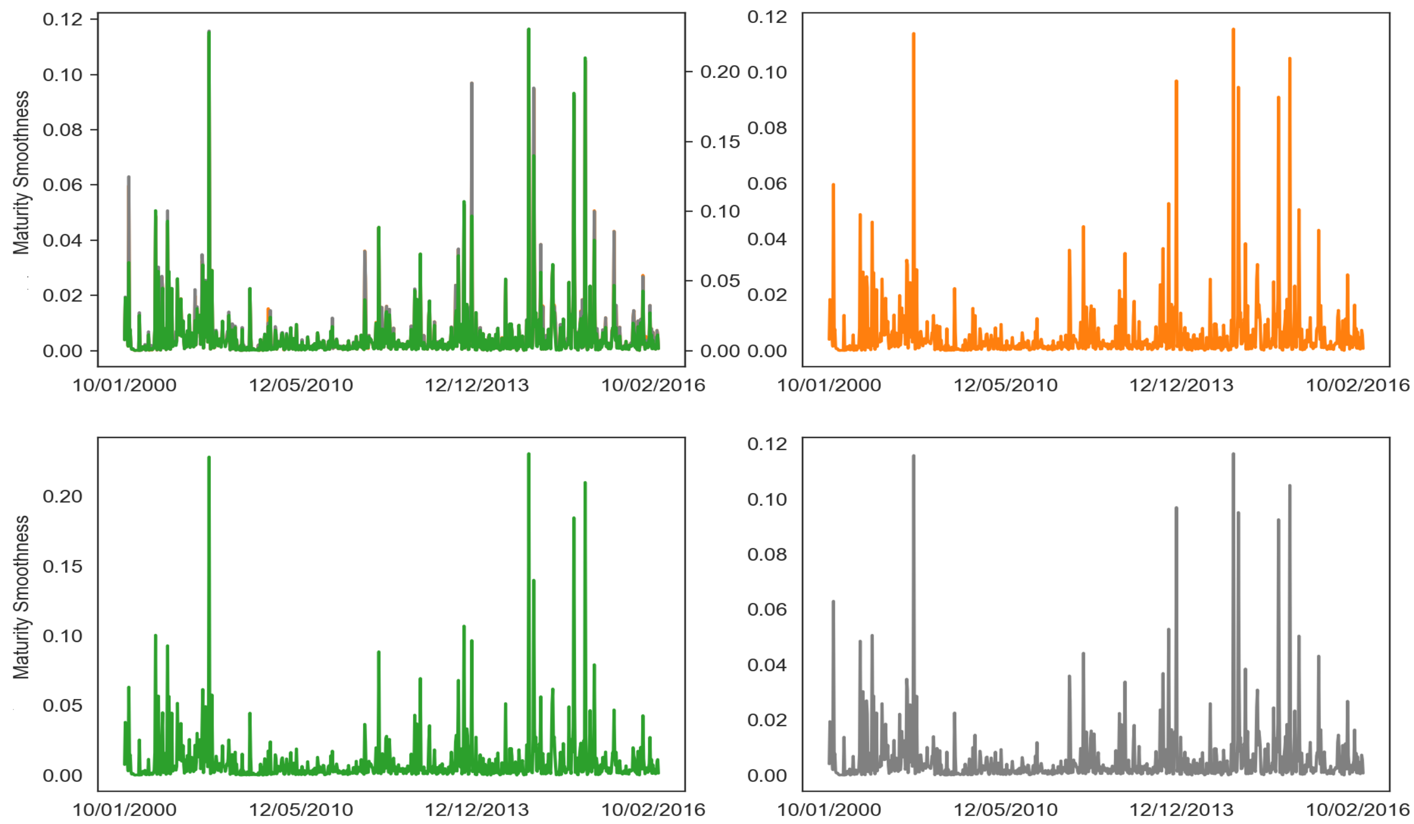
Note: this figure shows the relative distance of three methods. The top left panel plots all results. The green line represents Fengler and Hin (2015). The orange line represents Fengler (2009) and the gray line represents the result of L_1-SVM.

Figure 5.5.3: Forward-moneyness Smoothness Comparison



Note: this figure reports the smoothness(in forward-moneyness direction) of three methods. The top left panel plots all results. The green line represents Fengler and Hin (2015). The orange line represents Fengler (2009) and the gray line represents the result of L_1-SVM.

Figure 5.5.4: Maturity Smoothness Comparison



Note: this figure reports the smoothness(in maturity direction) of three methods. The top left panel plots all results in one figure. The green line represents Fengler and Hin (2015). The orange line represents Fengler (2009) and the gray line represents the result of L_1-SVM.

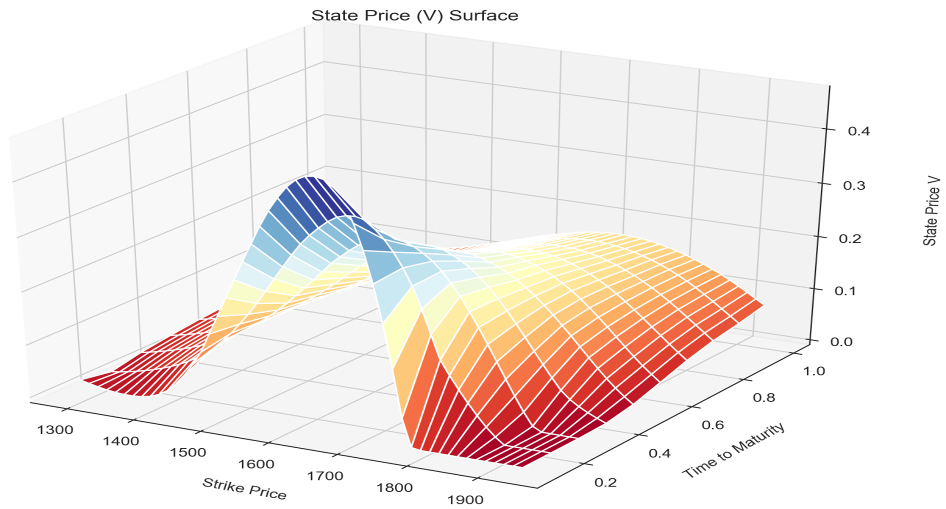
spread no-arbitrage condition is a weak constraint. Overall, my method displays a relatively smooth call price surface.

From the above analysis, my L_1-SVM method provides a universal framework that incorporates previous studies by using different kernel function in Equation(5.14). Previous studies with univariate and bivariate kernels can both be replicated in my linear programming framework. Although my method requires additional time to search the optimal parameters, as shown in Table 5.5.2, it improves the estimation accuracy and surface smoothness in the forward-moneyness direction. Therefore, in this chapter, I use the estimated call option price from L_1-SVM to extract the state price density (SPD). Since the above comparison is under the forward measure, based on Equation(5.6), I first transform the estimated option price back to risk neutral measure and then calculate SPD based on Equation(5.5). Figure 5.5.5 shows the extracted SPD on 02 July, 2013. As expected, the SPD is unimodal, smooth and positive. The SPD becomes small for far OTM and far ITM option.

5.6 Summary

In this chapter, I investigate the problem of estimating risk neutral information (SPD or RND) from option price. I find estimation of SPD from option prices faces five challenges: (a) in Breeden and Litzenberger (1978)'s estimation equation, the strike price is continuous while strike prices in real markets are discrete; (b) market data contain noise that may lead coarse and multimodal SPD; (c) theoretically SPD starts from 0 and extends to infinity while the market option data can only estimate SPD within certain bounds; (d) the estimated call option price surface should incorporate no-arbitrage constraints; (e) the estimation of

Figure 5.6.1: State Price Density on July 2, 2013



Note: Figure 5.5.5 plots the state price surface on 02/07/2013. The index price is 1614.08. The state price density obtained as second oder derivatives of call price function respect to strike price.

SPD surfers the 'curse of differentiation'.

After briefly review existing parametric and non-parametric methods, I find that none of the methods is superior and has successfully solved all five challenges. More specifically, the parametric method is not flexible to satisfied all no-arbitrage constraints and thus leads to under-fitting the real market data. The non-parametric method shows good performance in approximating the surface but is sensitive to the pre-determined parameters. Therefore, differing from these two methods, I propose a new machine learning approach to estimate the call option price surface. Compared with parametric and non-parametric methods, machine learning has two advantages. First, since it is a data-driven approach, machine learning exhibits good performance in solving constrained optimization problems. Second, it is not sensitive to pre-determined parameters because the optimal value of these parameters are chosen using the gridsearch technique during training.

Based on empirical studies, although most use the average of the bid and ask price as fair option price, the true transaction option price lies in a range. To take this into account, I develop a data-driven approach L_1-SVM based on standard support vector machine(SVM)[49], which incorporates the information in the bid-ask spread in pre-defining error tolerance in a loss function. As shown in Section 5.3.2, my L_1-SVM method is sufficiently flexible to consider different models and all arbitrage-free constraints.

Empirically comparing my L_1-SVM method with other non-parametric methods using S&P 500 index options, I show that my method is accurate and smooth. It is easy to implement and can be universally applied by choosing different kernel functions. Previous studies that use cubic spline, low-order polynomial and tensor product spline estimation method can all be replicated in my framework.

[49]See Section 3.1.2 for more detail.

Chapter 6

Extracting Natural Probabilities from Option Prices: The Empirical Ross Recovery Theorem

All models are wrong, but some are useful.

—Box (1976)

6.1 Introduction

Traditionally, practitioners and academics have been content with forecasting returns based on historical data, where future returns are assumed to be drawn from specific models or distributions. Although this standard econometrics approach yields relative a good performance in predicting stock volatility, it fails to capture extreme movements in the financial markets. History has clearly revealed that rare events in the financial markets occur with a low probability which, from a statistical perspective, lies in the tail of a probability distribution(Liu (2014)). When I use historical data to infer future returns, these extreme values are likely to be treated as outliers. Hence, valid inferences on future returns can only be obtained by incorporating the tail dependence(Poon et al. (2003)) and adding additional restrictions.

Is it possible that I can obtain a valid inference without using historical data? Researchers have investigated this issue in the derivatives market. With payoffs extending into future, financial derivatives naturally offer a prism through which ay market's perception of future returns can be revealed. The hope that this revealed information can be used to forecast underlying future returns has fascinated academics and practitioners. While there exists a rich equity option market and a literature about how to extract risk-neutral probabilities from option prices (see Chapter 5 and Anagnou et al. (2002) for a review), previous studies do not consider directly extracting real-world probabilities from observed option prices. This is for two reasons. First, classic option pricing theory is based on a risk-neutral valuation framework(Black and Scholes (1973); Merton (1973)), in which there is no role for the dynamics of the underlying under a real-world measure $\mathbb{P}$. If the option is valued without knowing the underlying dynamics under $\mathbb{P}$, how can this information be inferred from option prices? It seems impossible to recovery real-world probabilities from option prices. Second, as suggested by Jackwerth (2000) and shown in Section 5.2.1, in each state of the world, the relationship between risk neutral price and real-world probability and risk aversion is shown as below.

$$risk \quad neutral \quad probaility = stochastic \quad discount \quad factor \times real \quad world \quad probaility$$

Apparently, from this relationship, there is an infinite combination of stochastic discount factor and the real-world probability that yield the same risk neutral probability. Economically, this means when I extract the risk-neutral probabilities from the prices of options on the S&P 500, I find the risk-neutral probability of, for example, a 25% drop in one month to be higher than the probability calculated from historical stock returns. Since the risk-neutral probability is the real probability adjusted for the stochastic discount factor, this result indicates either that the market forecasts a higher probability of a stock decline (implied by real-world probability) than has occurred historically or the market requires a very high risk premium (implied by stochastic discount factor) to insure against a decline. Without knowing which is the case, it is impossible to separate the two and infer the market's forecast of the real-world probability. Therefore, in order to estimate real-world probabilities from option prices, the stochastic discount factor (or pricing kernel in asset pricing literature) should be known. However, there is no widely accepted measure of the stochastic discount factor in previous studies, this therefore leads to controversial recovered results.

Recently, without appealing to historical data and assuming an investor's risk preferences, Ross (2015) poses "The Recovery Theorem (TRT)", which enables the recovery of real-world probability and stochastic discount factor together from Arrow-Debreu state prices (Arrow (1964)) implied by observed option prices. This is a remarkable study and the result relies on two insights: (a) the stochastic discount factor is transition independent; (b) there is a time homogeneous, finite state Markov chain process driving all changes in the economy. These two assumptions together allow for a unique recovery of real-world probability.

Following Ross's approach, several studies attempt to generalize TRT by relaxing these two key assumptions. As pointed out by Huang and Shaliastovich (2014), current economic conditions may influence future wealth and, therefore, the stochastic discount factor implied utility in Ross (2015)'s first assumption is incorporated with Kreps and Porteus (1978)'s recursive preferences. By considering these, they extend TRT to an Epstein-Zin type utility based framework, in which when wealth to consumption ratio is given, recovered pricing kernel captures the investor's preference for timing uncertainty.

Additionally, several recent papers focus on extending the second assumption of TRT from discrete time to continuous and from bounded state space to unbounded. Replacing the discrete-time Markov chain with a continuous time recurrent Borel right process, Qin and Linetsky (2016) prove that TRT is possible when the stochastic discount factor is a positive semi-martingale function. This semi-martingale stochastic discount factor setting is further extended and linked to Hansen and Scheinkman (2009) by Qin and Linetsky (2017). Changing the risk neutral measure to T-forward measure, Qin and Linetsky (2017) extend TRT without the Markovian assumption. Clearly, TRT should extend to continuous time since the real financial market is in continuous time, but whether I need TRT in unbounded state space is debatable. Some researchers argue this is not necessary because even if the true economy state space is unbounded, I always truncate the space because I have a limited number of observations. However, I argue that the extension of TRT is required for two reasons. First, although there is no conclusion about whether the economy state pace is bounded or unbounded, for robustness, it is necessary to extend Ross (2015) to an unbounded diffusion setting. Second, even though I assume the state space is bounded, it is impossible to determine the upper bound of economy state space. Take the S&P 500 index as an example, no one can know a correct upper bound .

The study of the boundary condition of TRT is initiated by Dubynskiy and Goldstein (2013), who prove the boundaries of state space are essential for separating risk aversion and real-world probability. However, they argue the assumed boundaries which, in order to apply TRT, fundamentally change the solution of real-world probability. This result is confirmed by Walden (2017), who shows that boundaries only concern the state dynamics instead of the pricing kernel. In contrast to Dubynskiy and Goldstein (2013)'s finding of boundaries, by relaxing the bound restriction, Walden (2017) shows that recovery is possible in an unbounded diffusion-type state space when the diffusion process is recurrent and the sufficient but not necessary condition for deriving unbounded is a mean reverting diffusion process.

Instead of recovering from the Arrow-Debreu state prices perspective, Carr and Yu (2012) consider an alternative set of sufficient conditions to derive TRT. Relying on restricting the dynamics of the numeraire portfolio, they show the unique real-world probability can be recovered using the regular Sturm-Liouville theory. This study provides an insight into recovering real-world probability with a preference-free approach. Along the line of this perspective, Park (2016) further supports Walden (2017) by proving recovery under a recurrent and mean reverting diffusion process is possible under Carr and Yu (2012)'s framework.

Although previous studies have successfully extended TRT, Borovička et al. (2016) point out that there is a theoretical flaw in TRT, arguing that Ross' first assumption is a special case of Hansen and Scheinkman (2009), in which the martingale component in the stochastic discount factor factorization equals one. They suggest that this special case only holds in an economic environment where the highest return asset is the long bond. In particular, they prove that, under the general stochastic discount factor factorization, what is recovered via the Perron-Frobenius theory is a distorted probability since the martingale component serves as the change of measure. This distorted probability reveals the market's long-term perspective. This distorted probability is also discussed in Alvarez and Jermann (2005), Hansen and Scheinkman (2009) and Qin and Linetsky (2017).

Empirical studies implementing TRT in stock options and bond futures options markets show conflicting results. In the context of the fixed income market, Qin et al. (2016) and Bakshi et al. (2017) test the potential misspecification of

TRT in Borovička et al. (2016)'s long-term factorization framework. Qin et al. (2016) conduct a hypothesis test for Ross's first assumption on the US Treasury market. They show that the recovered measure is different from the estimated real-world probability measure. The difference between these measures can be explained by the instantaneous volatility of the martingale component. Similarly, using the 30-year Treasury bond futures option, Bakshi et al. (2017) demonstrate that the martingale component of the stochastic discount factor exhibits stochastic behavior. These two studies confirm Borovička et al. (2016)'s factorization of the stochastic discount factor and show Ross's assumption (the special case of martingale component equal to one) is violated in the long bond market[50] market. In the equity option market, using S&P 500 index option, Jackwerth and Menner (2017) empirically evaluate the short-term forecast ability of recovered probability. They show that statistically the 1-month realized returns are not drawn from recovered probability. However, using a neural network estimation framework, Audrino et al. (2015) show that the recovered distribution of S&P 500 index option contains predictive information. A trading strategy that uses recovered moment significantly outperforms one using this risk neutral information.

Focusing on two conflicting results in the equity option market, I argue there are two issues with previous studies. First, they both use discretization to implement TRT. As shown by Tran and Xia (2015) and Dubynskiy and Goldstein (2013), the discretized recovery result is sensitive to the number of states and bounds on the state space. As a result, without comparing the two methods in the same number of states and bounds on the state space, it is difficult to conclude whether or not the recovered result has predictive ability. Second, Jackwerth and Menner (2017) argue that the recovered result of Audrino et al. (2015) can be at-

[50]Long bond refers as zero coupon bond with long maturity.

tributed to the loss function (or penalty term) in the neural network, but, without directly theoretical proving and empirically testing, it is hard to argue whether the failure of TRT in Jackwerth and Menner (2017) is due to theoretical flaws in TRT or their implemented technique.

To provide empirical evidence for the recovery problem, following Walden (2017), this chapter implements TRT in a continuous time setting. To my knowledge, this is the first study to investigate the empirical validity of TRT in continuous time. Without incurring discretization error, this chapter makes two contributions. First, I complete Walden (2017)'s unbounded diffusion TRT framework by providing the empirical evidence. Using Audrino et al. (2015) 's quadratic loss function in my L_1-SVM framework, my study shows that using a quadratic loss function in state price estimation does not guarantee a well-recovered result. To further investigate whether the distorted recovery result is caused by the implementation procedure associated error or the flaws of TRT, I simulate option data using the Ornstein–Uhlenbeck (OU) process and define a specified m(x) function. In line with Jackwerth and Menner (2017) and my empirical study using the S&P 500, I find that the recovered probability is different from the OU process generated real-world probability. Second, I contribute to the empirical studies on multivariate TRT. Although Ross (2015) and Walden (2017) both show that TRT can be applied in univariate and multivariate settings, the empirical studies are scarce. Sanford (2017) is the only empirical study that tests Ross (2015)'s recovery theorem in a multivariate Markov chain. Including volatility and underlying price as state variable, they show that the forecast results from multivariate TRT are far superior to univariate TRT. However, this result does not directly prove that the recovered probability is the real-world probability. In this chapter, using the theoretical result from Walden (2017) Section 3.4, I directly calibrate the two-dimensional consumption model of Bansal and Yaron (2004). I confirm that the recovered probability is not the real-world probability.

The road map of this chapter is as follows. In Section 2, I provide a theoretical review of three key studies of TRT. Section 3 introduces the implementation procedure of TRT and how to back out key input parameters from option prices. I test the TRT in unbound diffusion using S&P 500 index option price and Ornstein–Uhlenbeck (OU) process generated synthetic data. Section 4 provides an example of bivariate TRT using Bansal and Yaron (2004)'s consumption model. In Section 5, I discuss the reasons that why TRT fails and Section 6 concludes.

6.2 The Recovery Theorem

The basic objective of TRT is to extract the real-world measure from the risk neutral measure. Until Ross (2015), this aim has always been viewed as impossible because it is hard to disentangle real-world probability from option implied information. As shown in Section 2.23 and Section 6.1, the risk neutral probability is a product of two components: (a) real-world probability, and (b) the stochastic discount factor. In general, option prices do not offer sufficient information to separate them. Surprisingly, Ross (2015) proposes TRT, which is a new way to recover the real-world probability from the risk neutral measure under certain assumptions. In this section, I provide a theoretical review of three key studies of TRT. I illustrate the basic assumptions and key recovery procedure of TRT under discrete and continuous time settings.

6.2.1 Discrete Recovery

6.2.1.1 Basic Framework

To successfully recovery the real-world probability, Ross (2015) makes two main assumptions. First, he assumes a discrete-time economy world, in which the driver of uncertainty follows a finite time homogeneous Markov chain process. This assumptions implies two essential characteristics for the uncertainty driver X: (a) X is bounded and lower and upper limits exist. (b) the future state of X only depends on the current state. Second, the stochastic discount factor (or pricing kernel) is transition-independent, in other words, the utility for a representative agent in the market is state-independent and additively separable. Using these two assumptions, Ross (2015) shows that the recovered real world probability problem transforms to an eigenvalue analysis of Arrow Debreu state prices.

Let us consider a economy world with a finite number of states M and time horizons N. The risk neutral information implied by Arrow Debreu state price S is defined as:

$$S = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,n} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m,1} & s_{m,2} & \cdots & s_{m,n} \end{bmatrix}$$

As the S is driven by a Markov chain process, I define the next state Arrow Debreu state price S_{n+1} only depending on current state S_n . I express this relationship as $S_{n+1} = \boldsymbol{Q} S_n$, where $\boldsymbol{Q}$ is an $M \times M$ transition matrix denoted by:

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,n} \\ q_{2,1} & s_{2,2} & \cdots & q_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m,1} & q_{m,2} & \cdots & q_{m,m} \end{bmatrix}$$

The transition matrix Q is the key to implementing Ross's Recovery. The rows of Q denote the current state and the columns represent the next potential state, so the sum of each row of Q will be one. If I define the real-world transition matrix as P, where

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,1} & p_{m,2} & \cdots & p_{m,m} \end{bmatrix}$$

My aim is to recovery P from Q, which is uniquely determined by Arrow Debreu state price S[51]. Under the general equilibrium pricing model (see Section 2.2.3), this two matrixes are connected with the pricing kernel. Assume the pricing kernel matrix is given as:

$$\Phi = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots & \varphi_{1,n} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots & \varphi_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m,1} & \varphi_{m,2} & \cdots & \varphi_{m,m} \end{bmatrix}$$

Using the transition independent assumption, the pricing kernel from state i to j can be written in the following form:

[51] Ross(2015) assumes the Arrow Debreu state price S is known. Please see Chapter 5 for the estimation of this matrix.

$$\Phi_{i,j} = \frac{q_{i,j}}{p_{i,j}} = \delta \frac{h(i)}{h(j)} \tag{6.1}$$

Where $h(.)$ is a positive function. δ is the time discount factor. If I define the diagonal matrix as

$$D = \begin{bmatrix} h(1) & 0 & \cdots & 0 \\ 0 & h(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h(3) \end{bmatrix}$$

Replacing $h(.)$ with D, then Equation(6.1) can be written as:

$$Q = P\Phi = \delta D^{-1} D \tag{6.2}$$

Rearranging this equation yields

$$\delta^{-1} Q D^{-1} = D^{-1} 1 \tag{6.3}$$

Setting $z = D^{-1}$, Ross(2015) further simplifies Equation(6.3) as an eigenvalue analysis problem

$$Qz = \delta z \tag{6.4}$$

Since the transition matrix Q is non-negative, applying the Perron-Frobenius Theorem(Meyer (2000)), Ross(2015) shows unique recovery of the real world probability (matrix F) and the pricing kernel (matrix D up to a positive scaling).

Theorem 6. *(Perron-Frobenius Theorem) If all elements in a square matrix A is a non-negative matrix, the following statements hold:*

1. A has a nonnegative real eigenvalue and the largest eigenvalue dominates the absolute values of all other eigenvalues.

2. the following problem has an unique pair of eigenvalue and eigenvector

$$Ax = \lambda x$$

Where x is a vector and λ is a scalar. The elements in the eigenvector are also strictly positive.

6.2.1.2 Discussion

While Ross(2015) elegant proves the recovery of the real-world probability from risk neutral measure is possible, I argue that the framework of TRT has two groups of limitations (see Table 6.2.1 for summary). In contrast with the most option pricing models, which are build on continuous time with a diffusion process, Ross (2015) assumes a discrete and bounded economy world. Undoubtedly, a continuous time setting may more appropriate as the real financial market is continuous. However, whether the state space must unbounded is unclear. In Ross's framework, the information of state bound is included in Arrow Debreu state price matrix S. The first row in S shows the information of low bound and the last row in Q presents the upper bound. As a result, the choice of bound is determined in the value of S, which further influences Q and the recovered results. This influence of boundaries is also mathematically proved by Dubynskiy and Goldstein (2013) and Tran and Xia (2015).

Moreover, Ross's assumption of investor's utility is also criticized by fellow researchers (Carr and Yu (2012); Borovička et al. (2016)) and challenged by empirical work on the stock market (Mehra and Prescott (1985)). In fact, the utility function itself is difficult to capture and measure. For example, how should I decide on the form of the utility function? According to the Expected Utility Theory, the utility function is expressed as probability weighted average of

Table 6.2.1: Limitations of Ross's Recovery Theorem

	Problem	Remark
Assumption Drawbacks	Discrete Time	Potentially problematic as the real world is in continuous time
	Boundaries, the best and worst state must be known	Walden (2017)argues, even if the true state space is bounded, it may even be unknown
	State independent and additively time separable utility	This assumption rules out habit formation and additive of utility may be applied with restriction
Implementation Restrictions	Can only forecast stock index, which is a proxy of entire holding of representative agent.	In other word, assets, which could not reflect holding of representative agent, cannot be forecast.
	Could not provide forecast for assets, whose underlying is zero net supply	This assumption is also related to representative agent, which requires underlying is positive net supply assets. Hence, assets such as Volatility Index (VIX) and future could not be forecast.
	Could not provide forecast for cash-settled derivative, whose underlying is not tradable. For example, underlying is temperature	Because the $h(.)$ function in Ross(2015) links to investor's marginal rates of consumption.

contingent utility. However, based on Prospect Theory, utility is a weighted related probability and a non-linear function, since people overweight small probability. The choice of utility function becomes extremely complex since Ross's assumptions involve time-additivity and transition independence. The transition independence of utility, means the utility of a representative agent only depends on final and current state and rules out both an economic satiation effect and habit persistence in utility function (Carr and Yu (2012)). The satiation effect, put simply, is the diminishing marginal utility effect. Habit persistence, or habit formation, captures the effect that an increase in current consumption increases its marginal utility in the next time period. since Ross's framework is build on the assumption that there exist a representative agent, the implementation of TRT is restricted. It can only be applied to forecasting an index asset because it is a proxy for the representative agent's holding. Besides, it could not be used to forecast the underlying, which is zero net supply and not tradable.

6.2.2 Continuous Recovery

To derive a general version of TRT and loosen the assumption of Ross, Carr and Yu (2012) and Walden (2017) demonstrate that similar separation can be achieved in a bounded and unbounded continuous time setting. In contrast to using the risk neutral measure as the Arrow Debreu state price, Carr and Yu (2012) accomplish the recovery task using the properties of a numeraire portfolio. As introduced by Long (1990), in a market with no arbitrage opportunities, there exists a portfolio under which the deflated asset price evolves at a constant expected rate. When examining Ross's assumption in the numeraire portfolio context, Carr and Yu (2012) replaces Ross's assumption about the utility with the restrictions on the driver X. They show that unique recovery is possible in a time-homogeneous diffusion setting when the boundary conditions are known.

6.2.2.1 Bounded Recovery

Consider an economy with probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is a unknown real world probability. Assume the risk-less asset (or money market account) is defined as

$$dS_{0t} = r_t S_{0t} dt \tag{6.5}$$

with a initial condition $S_{00} = 1$. There are no arbitrage opportunities in the market, therefore, according to the fundamental asset pricing theorem in Section 2.1, a unique risk neutral measure $\mathbb{Q}$ exists. Under $\mathbb{Q}$, the security price adjusted by risk free rate is a martingale

$$E^{\mathbb{Q}}[\frac{S_{iT}}{S_{0T}}] = \frac{S_{it}}{S_{0t}} \tag{6.6}$$

Using the numeraire portfolio maps the risk neutral measure $\mathbb{Q}$ to real-world measure $\mathbb{P}$, Carr and Yu (2012) denote the relationship between numeraire portfolio L_t and money market account as

$$L_t \equiv \frac{S_{0t}}{M_t} \tag{6.7}$$

Assume a univariate uncertainty process X exists in the economy and drives all the observables, such that,$r(x,t) \equiv r(x)$,$\sigma(x,t) \equiv \sigma(x)$. X satisfies the stochastic differential equation

$$dX_t = \beta(X_t)dt + \alpha(X_t)dW_t^{\mathbb{Q}} \tag{6.8}$$

Where $\beta(x)$ and $\alpha^2(x)$ are the known drift and variance function respectively, and $W_t^{\mathbb{Q}}$ is a Brownian motion under $\mathbb{Q}$. The infinitesimal generator $\mathcal{G}$ of X is defined by

$$\mathcal{G} = \frac{\partial}{\partial \mathrm{t}} + \frac{\alpha^2(x)}{2}\frac{\partial^2}{\partial x^2} + \beta(x)\frac{\partial}{\partial x} \tag{6.9}$$

Directly apply the Girsanov's Theorem in Section 2.1.2 to change the measure of X_t, the dynamics of X_t under $\mathbb{P}$ is

$$dX_t = (\beta(X_t) + \sigma_t(X_t)\alpha(X_t))dt + \alpha(X_t)dW_t^{\mathbb{P}} \tag{6.10}$$

Based on Equation(6.7), the $\mathbb{P}$ dynamics of the numeraire portfolio L_t can be derived by Ito's lemma, which yields

$$\frac{dL_t}{L_t} = (r(X_t) + \sigma^2(X_t))dt + \sigma(X_t)dW_t^{\mathbb{P}} \tag{6.11}$$

Where r_t is the short rate,σ_t^2 is the instantaneous variance and $W_t^{\mathbb{P}}$ is a Brownian motion under $\mathbb{P}$. From Equation(6.11), it is clear that the risk premium of the numeraire portfolio L_t is determined by its variance σ_t^2. In other words, if I can determine the volatility process of L_t, I can recover the $\mathbb{P}$ dynamics of the numeraire portfolio. If I apply Ito's formula to $L_t \equiv L(X_t, t)$ and substitute Equation(6.9), then the volatility of L_t is

$$\sigma(X) = \alpha(X)\frac{\partial}{\partial x} lnL(X,t) \tag{6.12}$$

Rearrange Equation(6.12) and take the exponentiation for both sides, the value of numeraire portfolio can be expressed by two separated terms

$$L(X,t) = \rho(X)\gamma(t) \tag{6.13}$$

Where $\rho(x) = e^{\int_l^x \frac{\sigma(y)}{a(y)}dy}$ and $\gamma(t) = e^{f(t)}$, $f(t)$ is the constant of integration and

$\rho(.)$ and $\gamma(.)$ are positive function. Substituting Equation(6.12) into the diffusion generator (Equation(6.9)), it implies

$$\frac{\alpha^2(x)}{2}\frac{\rho^{''}(x)}{\rho(x)} + \beta(x)\frac{\rho^{'}(x)}{\rho(x)} - r(x) = \frac{\gamma^{'}(t)}{\gamma(t)} \tag{6.14}$$

This equation can hold only if both sides are constant and so finding the unique solution of this problem fits into the the regular Sturm Liouville theorem. The general solution of this eigenfunction analysis problem yields the value of the numeraire portfolio having the form

$$L(X,t) = \rho(X)e^{rt} \tag{6.15}$$

Hence, imposing the restrictions on the $\mathbb{P}$ dynamics of the numeraire portfolio, Carr and Yu (2012) provide an alternative way to extract the real world measure. I argue that the restrictions of numeraire portfolio in Carr and Yu (2012) fundamentally connect to the utility assumptions of Ross (2015). As suggested by Kelly (1956) and Platen (2006), the numeraire portfolio has the optimal growth rate, which means the investors in the portfolio aim to maximize the logarithm of their terminal wealth. This is consistent with Ross (2015)'s framework, in which the representative agent aims to maximize his utility. However, the empirical implication of Carr and Yu (2012) is difficult because the numeraire portfolio is either not actually traded nor can it be constructed in the real market.

6.2.2.2 Unbounded Recovery

Inspired by Carr and Yu (2012), Walden (2017) further extends the recovery into the unbounded diffusion. Instead of restricting the numeraire portfolio, Walden (2017) extends Ross (2015)'s TRT framework on Arrow Debreu securities. In this section, I summarize the Walden (2017) approach and show how it is related to Ross (2015).

Similar to Carr and Yu (2012), Walden (2017) considers an economy with probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is a unknown real world probability. Suppose there is a state variable X that drives all the observations

$$dX_t = \mu(X_t)dt + \sigma(X_t)dw \tag{6.16}$$

Mathematically, to ensure there are strong solutions in any interval of Equation(6.16), Walden (2017) assumes that μ and σ are continuously differentiable. For any x and y, there are exists constant C_1,C_2 and C_3, such that $\frac{|\mu(x)-\mu(y)|}{|x-y|} \leq C_1$, $\frac{|\sigma(x)-\sigma(y)|}{|x-y|} \leq C_3$ and $0 < C_2 < \sigma(x)$.

The transition density function $f^t(x,y) = \frac{\partial F^t}{\partial y}$ associated with the above diffusion satisfies the Fokker-Planck equation, therefore

$$\frac{\partial f^t}{\partial t} = \mathcal{L} f^t \tag{6.17}$$

$$f^0(x,y) = \sigma_x(y) \tag{6.18}$$

Similar to Equation(5.3), in this framework, the price of call option $C^t(K)$ is written as

$$C^t(K) = \int_K^\infty (y-K)\frac{m(y)}{m(x)} f^t(x,y)dy \tag{6.19}$$

Based on Equation(6.16), the Arrow Debreu state prices that inferred from call option prices $C^t(K)$ are defined as

$$p^t(x,y) = e^{-\rho t}\frac{m(y)}{m(x)} f^t(x,y) \tag{6.20}$$

This definition of Arrow Debreu state prices corresponds with Equation(6.1) in

Ross's framework. In the terminology of Ross (2015), the pricing kernel is transition independent. For convenience, I define the function $q(x) = \frac{m^{'}(x)}{m(x)}$ and $z(x) = \frac{1}{m(x)}$. Clearly, $q(x)$ is related to the representative agent's utility in Ross's framework.

Similar to Ross (2015), I assume the Arrow Debreu security price is known, which means in Equation(6.20) $p^t(x, y)$ is known. As proven by Walden (2017), I can infer the underlying parameters ρ, $\mu(x)$, $\sigma(x)$ and $m(x)$ from $p^t(x, y)$. More formally, the recovery problem of unbounded diffusion process becomes one of solving the fundamental ODE

$$z'' + \frac{k}{D}z' + \frac{\lambda - r}{D}z = 0 \tag{6.21}$$

Where $D = \frac{\sigma^2(x)}{2}$,$z' = \frac{-q(x)}{m(x)}$ and $z'' = -\frac{1}{m(x)}(q(x)' - q(x)^2)$. r is the short risk-free rate. $\lambda = \rho$. $k(x) = \mu(x) + 2q(x)D(x)$. In fact, Equation (6.21) is similar to the eigenvector formulation of Ross (2015). Although I can recover $m(x)$ using Equation (6.21), more conditions are needed to guarantee uniqueness. Consistent with Carr and Yu (2012), Walden (2017) finds that the recovery is only related to the dynamics of the state variable.

Proposition 3. *The necessary and sufficient conditions for recovery of m(x) are*

$$\int_{-\infty}^{0} e^{-\int_0^x \frac{\mu(s)}{D(s)}ds} dx = \infty$$

$$\int_{0}^{\infty} e^{-\int_0^x \frac{\mu(s)}{D(s)}ds} dx = \infty$$

Overall, as shown above, Walden (2017) provides a general version of Ross (2015) in an unbounded continuous setting. With restrictions on the behavior of the state variable, he shows the recovery is possible in unbounded diffusion.

6.3 Empirically Recovery in Continuous Time

Following Walden (2017), I empirically investigate TRT in unbounded diffusion setting. The procedure for applying the TRT is shown in Figure 6.3.1. The first step for TRT in continuous time is extract state price from observed market price. In this chapter, I use L_1-SVM state price density estimator to extract state price. Compare with other parametric or nonparametric methods that are reviewed in Chapter 5, my estimator shows two advantages. First, as shown in Chapter 5, my L_1-SVM estimator is comparatively accurate and smooth. Second, applying my L_1-SVM estimator will allow us to examine the issues whether the quadratic penalty term (or loss function) in machine learning framework will influence recovered result. To apply the quadratic loss function, I modify the Equation(5.21) as

$$min_{(\alpha,b)}\frac{1}{2}||\alpha_i||^2 + C\sum_{i=1}^{N}|f(k,\tau) - C(k,\tau)| \tag{6.22}$$

According to Section 3.1.2, this problem can be transformed to a dual optimization problem

$$max_{\{\alpha,\alpha^*\}} - \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(X_i, X_j) - \varepsilon\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{n}Y_i(\alpha_i - \alpha_i^*) \tag{6.23}$$

subject to

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$$

Following Audrino et al. (2015) and use Equation(6.23), I remove the estimated results that violate three no arbitrage conditions (Equation(2.27), Equation(2.31) and Equation(2.32)) during modeling. I suggest this standard support

Figure 6.3.1: The Recovery Procedure

Note: This figure gives the procedure for applying TRT in unbounded diffusion in practice. The S represents the state price. r is the short risk-free rate. k and D are defined in Section 6.2.2.2

vector machine can be estimated using Python's scikit-learn library. The second step of recovery problem is back out r, k and D from option prices, which has illustrated by Walden (2017) as follow:

Proposition 4. *For all $t \in (0, T)$, If the second order derivative of call option prices is defined as $V(t, y) = p^t(x_0, y)$. Then, for each y and t>0, this second order derivative has following relationship*

$$V_t = D(y)V_{yy} + \alpha_1(y)V_y + \alpha_0(y)V$$

where

$$\alpha_1(y) = 2D^{'}(y) - k(y)$$

$$\alpha_0(y) = D^{''}(y) - k^{'}(y) - r(y)$$

Where V_t,V_{yy} and V_y are the partial derivatives of V,D and k are defined in Section 6.2.2.2. $r(y) = \rho - q(y)\mu(y) - (q^{'}(y) + q(y))D(y)$ and $q(x) = \frac{m^{'}(x)}{m(x)}$

Please refer Walden (2017) Section 3.8 for proof

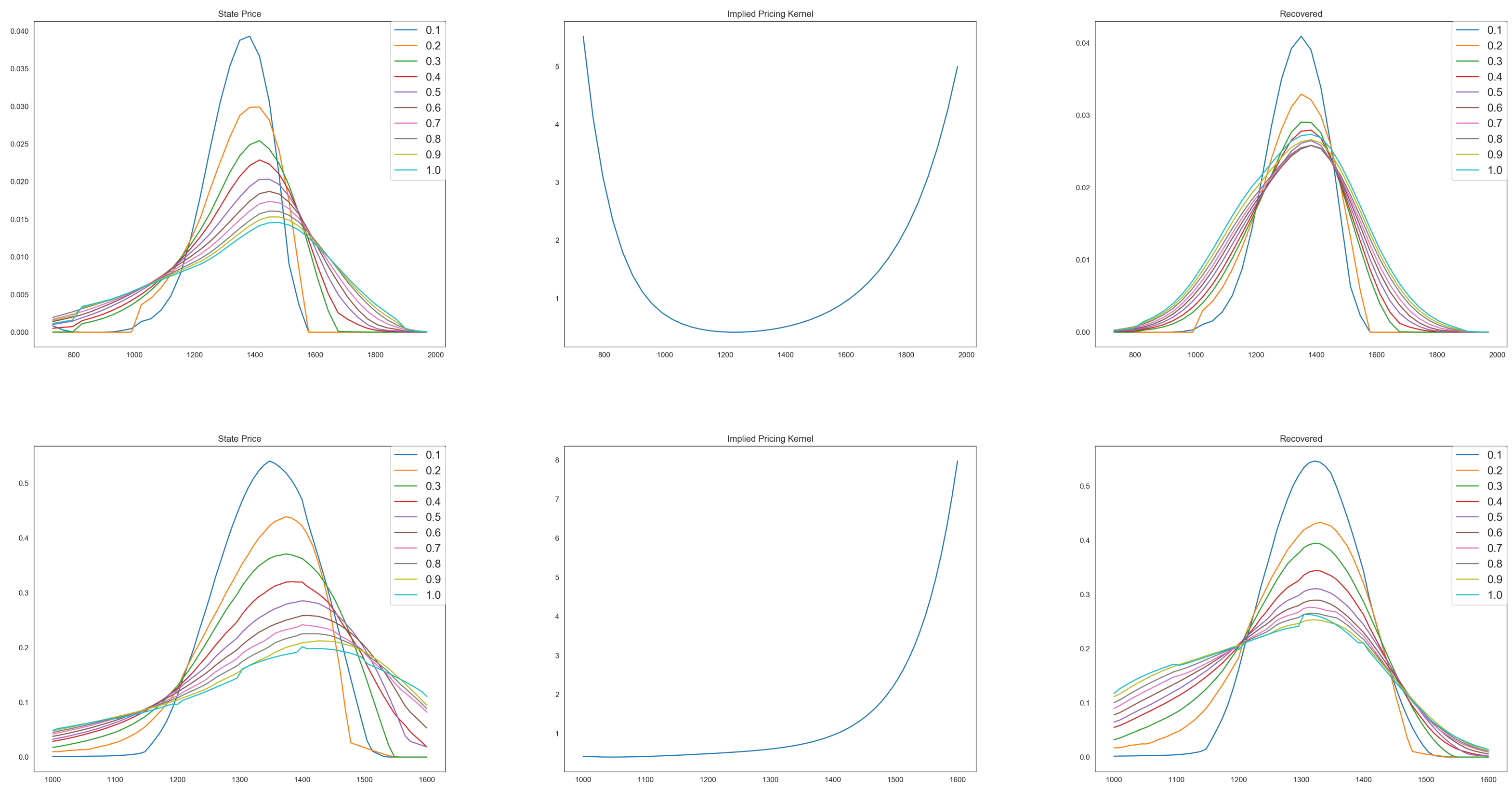
6.3.1 S&P 500 Index Option Example

In this chapter, I estimate the state price using the same dataset with Chapter 5. The dataset includes daily observations of S&P 500 index option from January 5, 2000 to April 30, 2016. I only use the OTM option data and apply the three level data filter proposed in Chapter 4. I estimate the state price using a standard quadratic programming support vector machine (Equation(6.23)). The description of the whole dataset is shown in Table 5.4.1. In Figure 6.3.2, I report two recovered results. The left panels show the state prices estimated from Equation(6.23). The right panels show the recovered physical distributions and middle panels report corresponding pricing kernels. Theoretically, the pricing kernel of S&P 500 index should monotonically decrease with state variable, in my case, the forward-moneyness. However, the empirical studies (such as Jackwerth (2000), Rosenberg and Engle (2002) and Constantinides et al. (2013)) show that the pricing kernel may increase locally. In my results, as shown in Figure 6.3.2, I do not find a persistent pricing shape kernel. Most surprisingly, instead of decreasing with forward-moneyness, the pricing kernel of bottom panel in Figure 6.3.2 even increase with it.

Specifically, to compare my continuous recovery with Ross (2015), in Table 6.3.1, I present the descriptive statistics of S&P 500 Options on April 27, 2011[52]. There are total 62 observations and the time to maturity from 7 to 365 days. Figure 6.3.3 illustrates the estimation results from Equation(6.23). Consistent with Equation(2.31), the call option surface is decreasing across the strike price and the state price is unimodal, smooth and positive. Figure 6.3.4 shows the corresponding partial derivatives of state price in Proposition 4. The recovered pricing kernel and probability are reported in Figure 6.3.5.

[52]This is the same date utilized in Ross (2015).

Figure 6.3.2: State Price, Recovered Pricing Kernel and Probability



Note: The first columns compares the recovered state price from Equation(6.23). The second columns illustrates the recovered pricing kernels, which is defined as the ratio of the state price to the real world probability. In each subplot, selected time maturities (0.1 to 1.0) are indicated by different color. The third column shows the recovered probability density. The top panel shows the recovered results on 05/03/2008 and the bottom panel reports the recovered results on 12/03/2011.

Table 6.3.1: Descriptive Statistics of S&P 500 Options on April 27, 2011

Panel A: Descriptive Statistics

				Percentiles					
Variables	Mean	Std. Dev.	Min	5%	10%	50%	90%	95%	Max
Call Price ($)	13.31	15.13	0.08	0.28	0.50	8.15	33.90	43.75	146.3
Put Price ($)	12.77	14.31	0.08	0.50	0.93	7.60	32.20	42.05	151.50
Implied Volatility σ (%)	20.62	8.39	6.46	10.30	11.46	19.19	31.10	35.69	87.11
Strike price(K)	1289.51	103.35	1100	930	1030	1540	2030	2110	1475
Days(τ)	53.82	41.47	7	14	18	44	95	127	365
Index Price (S)	1601.16	389.84	676.53	944.89	1091.60	1606.28	2079.51	2098.48	2130.82
Trading Volume (V)	1321.35	3690.84	1	2	4	110	3581	6698	157542

Note: Table 6.1 summaries descriptive statistics of S&P 500 Options on April 27, 2011. The total number of observations is 62 .The call price is calculated as average of bid and ask price. The OTM put prices are translated into ITM call prices by put and call parity. Std.Dev denotes the standard deviation from call option price.

In contrast to Figure 11 in Ross (2015), I do not find a strictly decreasing pricing kernel. My recovered pricing kernel exhibits a U shape, which provides new evidence to Rosenberg and Engle (2002) using a forward-looking information. Overall, I conclude I can recover the probability using Walden (2017) but what information do I recovery is not clear.

To test whether the recovered probability is drawn from real-world probability. Following Walden (2017), I test following null hypothesis using Berkowitz (2001) test, which has better accuracy in small samples Bliss and Panigirtzoglou (2004).

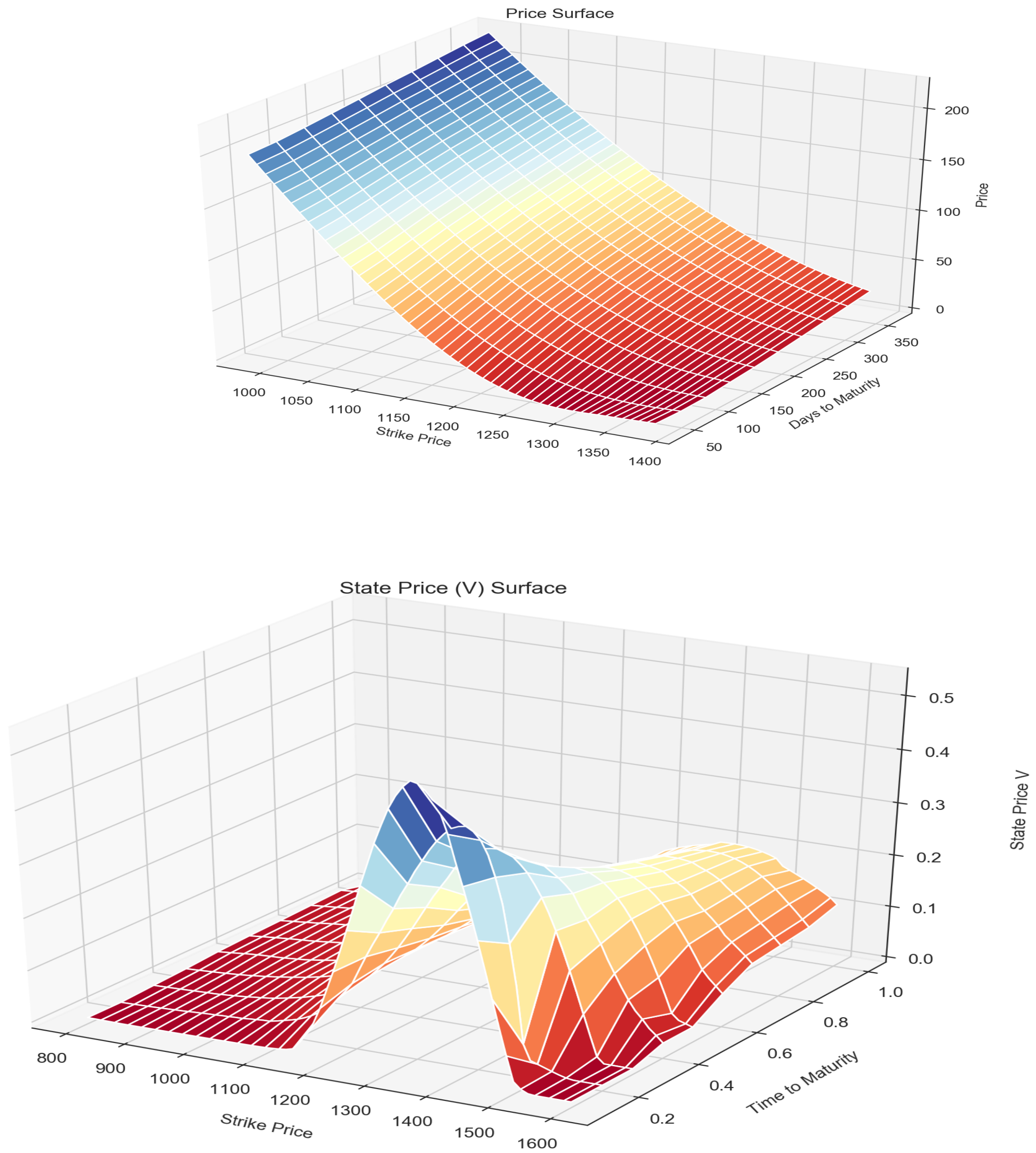
H_0 : Future returns are drawn from the recovered probability distribution.

In particular, I test this hypothesis with a monthly time horizon. I test whether the 1-month future realized S&P 500 index return is drawn from its real-world counterparts. If null hypothesis holds, I concludes that the recovered probability is the real-world probability. The Recovery Theorem is well specified or defined. The test procedure can be summarized as follow.

- First, I estimate the recovered probability p_i following procedure in Figure 6.3.1.
- Second, for each option, I find all 1-month horizon[53] options and their start and expiration dates.
- Third, I record the return between these start and expiration dates as 1-month realized return R_i.
- Finally, test the hypothesis whether R_i is drawn from p_i using Berkowitz test.

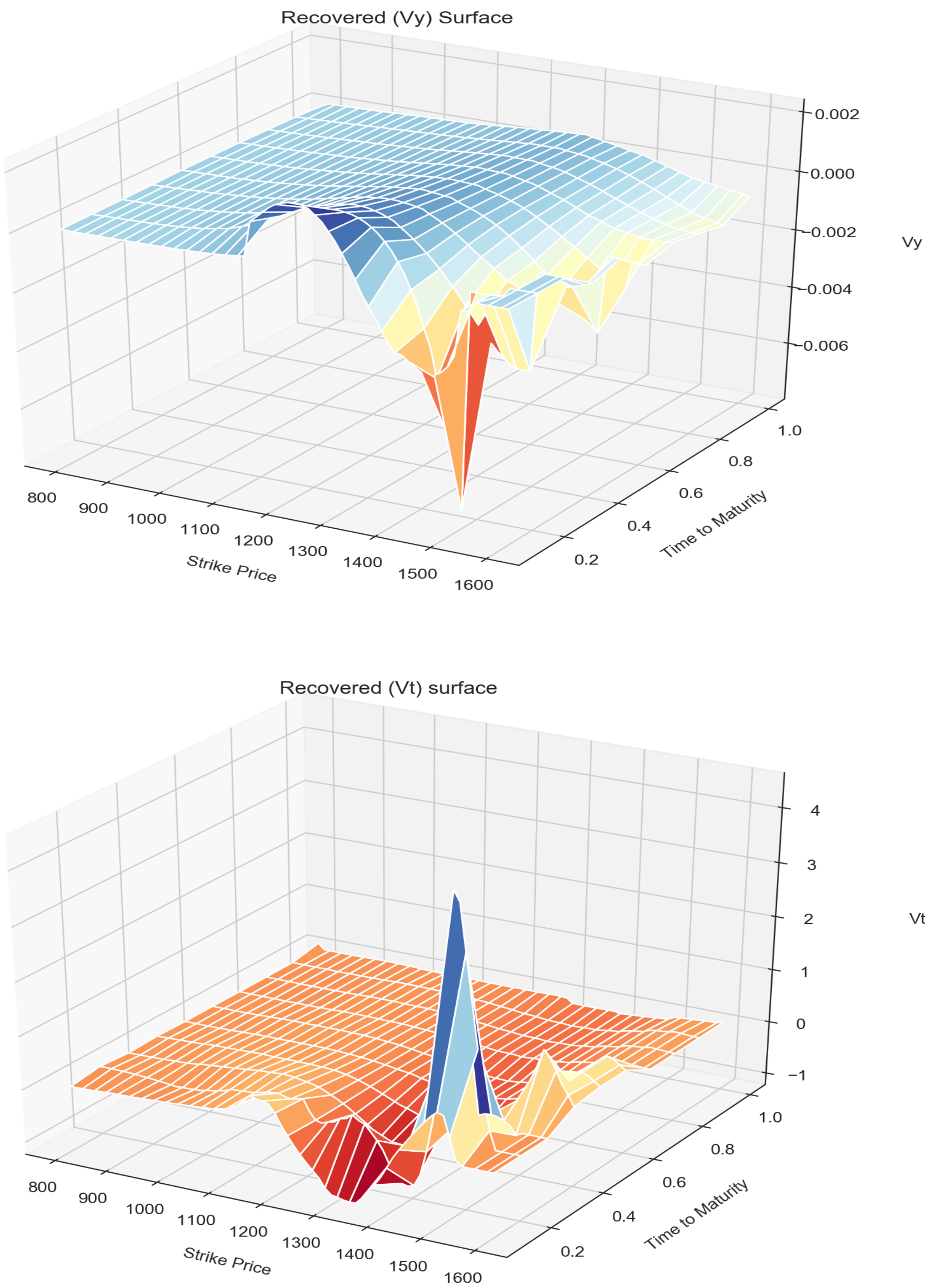
[53]This means the days to maturity is 30. In my sample, I accept days to maturity in[21,33] as 1-month horizon.

Figure 6.3.3: Estimated Price and State Price Surface on April 27, 2011



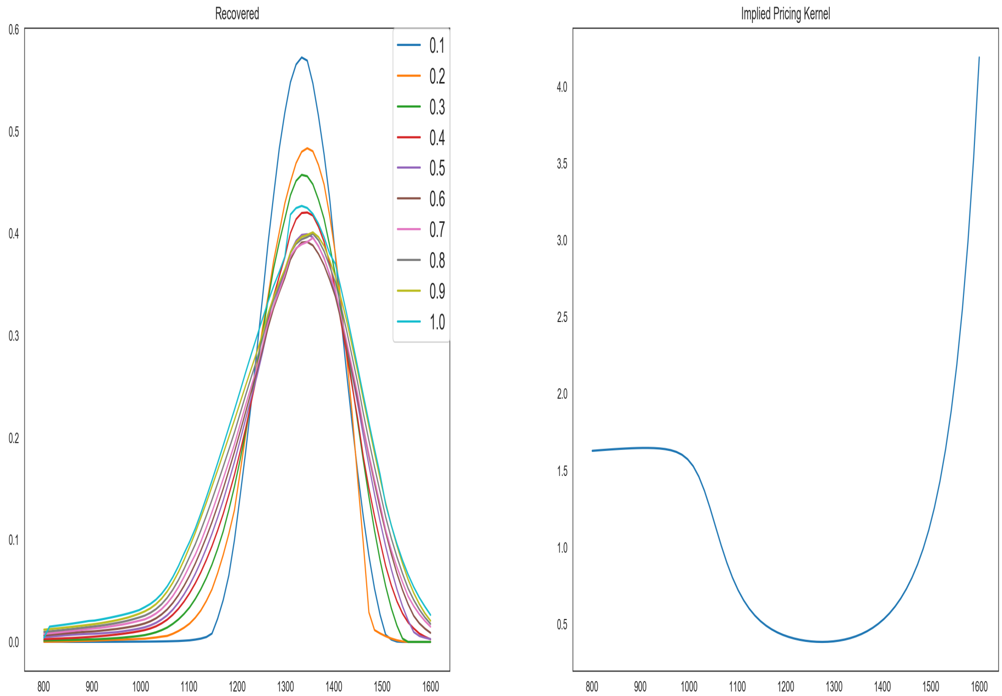
Note: Figure 6.3.4 shows the call option price and state price surface and on 27/04/2011. The top panel shows the call option price surface. The bottom panel displays the state price surface.

Figure 6.3.4: Partial Derivative of State Price Surface on April 27, 2011



Note: Figure 6.3.4 shows the associated partial derivative of state price surface on 27/04/2011.The top panel plots its first order derivative respect to the strike price. The bottom panel displays the first-order derivative of state price surface with respect to time to maturity.

Figure 6.3.5: Recovered Probability and Pricing Kernel on April 27, 2011



Note: Figure 6.3.5 plots the recovered results of S&P 500 index option on April 27, 2011. The right panel shows the recovered pricing kernel and the left panel reports the recovered probability density.

Table 6.3.3: Test Result of Recovered Probability

Period	Berkowitz(p-value)
2000-2007	0.001
2008-2009	0.005
2010-2016	0.001

Note: This table present test results for whether the 1-month realized return are drawn from recovered probability.

The test results are reported in Table 6.4. Similar to Walden (2017), my results strongly reject the null hypothesis.

6.3.2 Ornstein–Uhlenbeck Example

To investigate whether the distorted recovery result is caused by error associated with implementation procedure or comes from a theoretical flaw in TRT then, following Walden (2017), I consider the state evolving with the Ornstein–Uhlenbeck (OU) process

$$dX = \theta(a - X)dt + \sigma dw \tag{6.24}$$

where θ and σ are greater than zero. Since the OU process tends to drift towards its mean, it is generally used to model the mean reverting behavior of underlying in mathematical finance. As I can always define a new state variable $\dot{X} = X - a$, assuming $a = 0$, I further simplify Equation (6.24) as follow

$$dX = \theta(-X)dt + \sigma dw \tag{6.25}$$

Comparing above equation with Equation (6.16), I define $\mu = -\theta x$. According to Proposition 3, I check whether recovery is possible for this diffusion. Since $\frac{\mu}{D} = -\frac{\theta}{\sigma}x$, then I get $e^{-\int_0^x \frac{\mu(s)}{D(s)}ds} = e^{\int_0^x \frac{\theta}{\sigma}sds} = e^{\frac{\theta}{\sigma}x}$. Therefore, two conditions in Proposition 3 are satisfied. The recovery is possible for this unbounded diffusion.

To test whether the recovered result is affected by the estimation of Arrow Debreu state prices, in this section, instead of extracted state prices, I use simulated state prices based on Equation (6.20) directly. The empirical procedure can be summarized as follows.

- Generate the OU probability density under $\mathbb{P}$ measure by assuming $\mu = 1.1$, $\sigma = 0.3$, $\theta = 1$ and the initial underlying price $S_0 = x = 1900$.

- Get Arrow Debreu state prices by assuming $m(x) = 1 + x^2$ and using Equation(6.20).
- Back out r, k and D from the Arrow Debreu state prices
- Obtain the recovered probability

In the top panel of Figure 6.3.6, I plot my simulated option prices. The x-axis is the strike price and the y-axis is the option price. The strike price starts from 50 to 4000 and time to maturity is from 0.1 to 0.8. There are total 5400 option prices in my sample. Consistent with the real market option data plotted in Figure 5.4.1, the option prices reflect the decreasing behavior as I expect. The ITM option is more expensive than OTM option. Using the definition of r and k in Proposition 3, I obtain

$$r = \rho + \frac{2x^2\theta - \sigma^2}{1 + x^2} \tag{6.26}$$

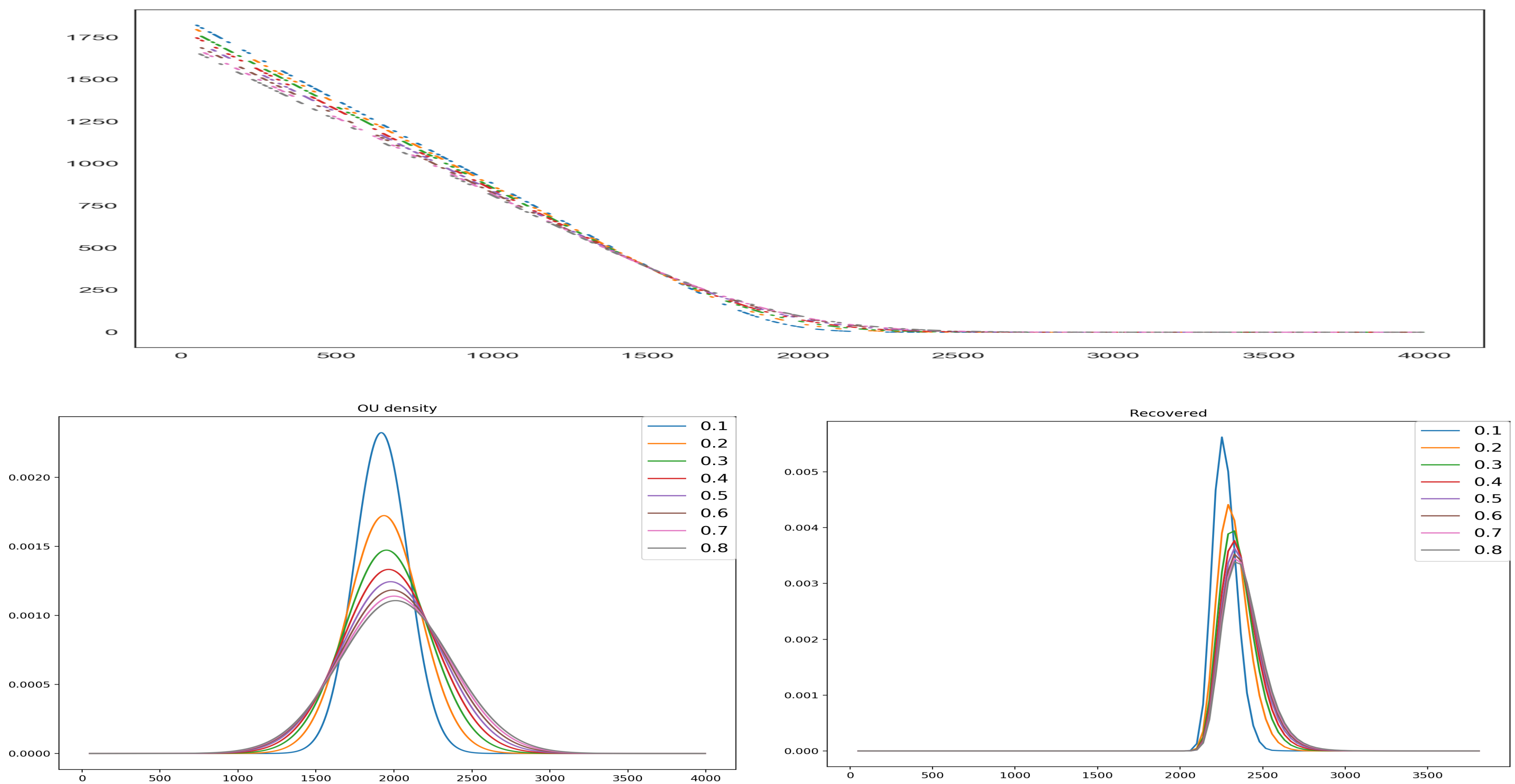
$$k = -x\theta + \frac{2x\sigma^2}{1 + x^2} \tag{6.27}$$

The fundamental ODE problem Equation(6.21) becomes

$$z'' + \frac{x}{D}(\frac{2\sigma^2}{1 + x^2} - \theta)z' + \frac{1}{D}(\lambda - \rho - \frac{2x^2\theta - \sigma^2}{1 + x^2})z = 0 \tag{6.28}$$

Solving this equation, I plot the recovered probability in bottom panel of Figure 6.3.6. I can see that the recovered probability density is significantly different from the OU density distribution under $\mathbb{P}$. The recovered distribution exhibits fatter tails than the real world distribution. Except for the options with 0.1 time to maturity, the recovered distributions with other times to maturity

Figure 6.3.6: Results for Ornstein–Uhlenbeck Process



Note: this figure plots the simulated data and recovered result for Ornstein–Uhlenbeck process. The top panel plots the simulated call option prices. There are total 5400 option prices in my dataset. The bottom panel shows the recovered density and the real OU density under P measure. The index price is 1900.

are significantly higher than the real-world probability. Consequently, instead of associated state price estimation error, the distorted recovery result is caused by a theoretical flaw in TRT.

6.4 Reasons for Distorted Recovery

6.4.1 Theoretical Explanation

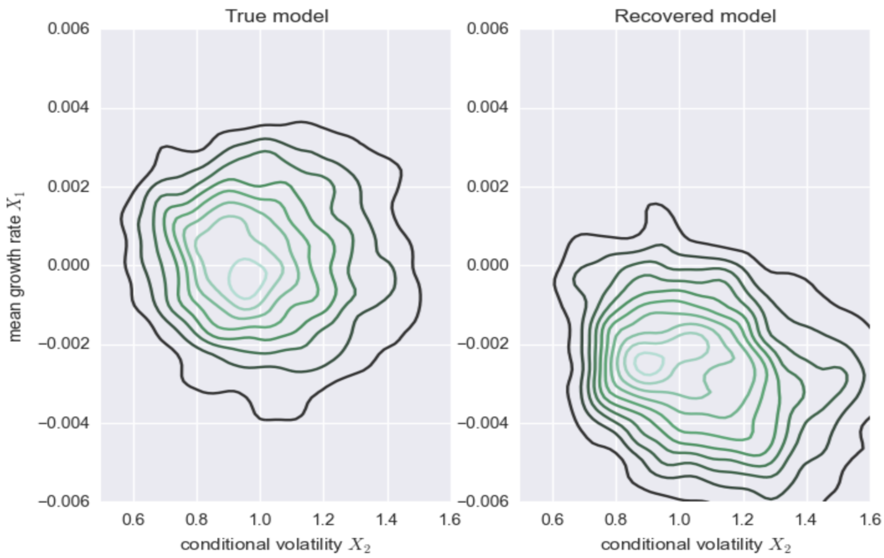
As criticized by Borovička et al. (2016), Ross's Recovery Theorem would lead a misspecified pricing kernel, which further results in a distorted real-world probability. They demonstrate that although Ross's first assumption (transition independence of pricing kernel) mathematically delivers a unique recovery, it is inconsistent with stochastic discount factor decomposition of Alvarez and Jermann (2005). The transition independence assumption implies that there is a martingale component associated with the stochastic discount factor and the recovered probability from Ross (2015) actually reflects a long term expectation.

To illustrate this issue, let us consider a discrete time environment with probability space $(\Omega, \mathcal{F}, \mathbb{P})$. There exists a state process X_t that is strictly stationary and time-homogenous. If I define the stochastic discount factor process as S_t, then based on Equation(2.2.2), at each time t, the price of claims that receive payoff Z_{t+n} at t+n, is given by

$$E[\frac{S_{t+n}}{S_t} Z_{t+n} | X_t] \tag{6.29}$$

According to Hansen and Scheinkman (2009), if I further restrict $Z_{t+n} =$

Figure 6.4.1: Recovery for Bivariate State Variables



Note: This figure shows the true and recovered joint distribution of state vector (mean growth rate and stochastic volatility). The parameters of the model is $a = -0.021$, $\bar{\sigma_{X1}} = [0 \quad 0.00034 \quad 0]'$ and $\bar{\sigma_{X2}} = [0 \quad 0 \quad -0.038]'$

$\psi(X_{t+n})$ and introduce a collection of linear pricing operators as $\{M_n\}$, I can get

$$M_n\psi(x) = E[\frac{S_{t+n}}{S_t}\psi(X_{t+n})] \tag{6.30}$$

Using the Perron-Frobenius Theorem and defining Φ is the positive eigenfunction/eigenvector and ρ is associated eigenvalue, I can further simplify the pricing operator M_n as[54]

$$M_n\phi = \rho^n\phi \tag{6.31}$$

This means I that can decompose the pricing kernel into two parts, namely permanent component $M_t^P = \rho^{-t}M_t\phi(X_t)$ and transitory component $M_t^T = \rho^t\phi(X_t)$(Alvarez and Jermann (2005) and Hansen and Scheinkman (2009)). There-

[54]Please see Hansen and Scheinkman (2009) and Hansen (2013) for more detail.

fore the SDF can be expressed as

$$\Phi = \frac{S_{t+n}}{S_t} = \frac{M^P_{t+n}}{M^P_t}\frac{M^T_{t+n}}{M^T_t} \tag{6.32}$$

where

$$\frac{M^P_{t+n}}{M^P_t} = \rho^{-1}\frac{M_{t+n}}{M_t}\frac{\phi(X_{t+n})}{\phi(X_t)} \tag{6.33}$$

$$\frac{M^T_{t+n}}{M^T_t} = \rho\frac{\phi(X_t)}{\phi(X_{t+n})} \tag{6.34}$$

If I compare Equation (6.32) with Equation (6.1), apparently Ross (2015) assumes the transitory component [Equation (6.34)) is equal to 1]. This is an unrealistic assumption Alvarez and Jermann (2005) and means the asset that can achieve the highest return is the long bond. Hansen and Scheinkman (2009) further shows that this transitory component is actually a martingale component, which defines a new probability that reflects long-term risk adjustments. As a result, due this martingale component, using Ross's TRT, I always get a distorted result.

6.4.2 Martingale Extraction using Bivariate Recovery

To test whether there is a martingale component in the pricing kernel, in this section, I apply Walden (2017) to Bansal and Yaron (2004)'s two-dimensional consumption dynamic model. Assume the state variables X_1 and X_2 evolve according to

$$dX_1 = -aX_2dt + \sigma_{X_1}dw_{X_1} \tag{6.35}$$

$$dX_2 = -X_1dt + \sigma_{X_2}w_{X_2} \tag{6.36}$$

where X_1 is growth rate and X_2 is stochastic volatility, w_{X_1} and w_{X_2} are independent Brownian motion processes. Based on Equations (6.33) and (6.20), I express the pricing kernel as $\Lambda_t = e^{-\rho t} \frac{m(X_t, Y_t)}{m(X_0, Y_0)}$. If I assume the aggregate consumption $C_t = e^{X_t}$, then the pricing kernel can be expressed as $\Lambda_t = e^{\rho t} \frac{e^{-rX_t}}{e^{-rY_t}}$, where γ is the risk aversion coefficient for a representative agent with power utility. Using this assumption and based on Equation(6.20), the recovery problem becomes one of solving the following PDE

$$\frac{\sigma_{X_1}^2}{2} z_{X_1 X_1} + \frac{\sigma_{X_2}^2}{2} z_{X_2 X_2} - X_2 z_{X_2} - (\alpha X_2 + \gamma \sigma_{X_1}^2) z_{X_1} + (\lambda - \rho + \gamma \alpha X_2 + \gamma^2 \frac{\sigma_{X_1}^2}{2}) z = 0 \tag{6.37}$$

In Section 3.4 of Walden (2017), he mathematically proves that this PDF has a unique positive solution. Thus, I can directly recover X_1 and X_2. The solutions of this PDE are related to discrete martingale component extraction in Borovička et al. (2016) Section 5. Following Borovička et al. (2016) and Hansen et al. (2007), if the recovered X_1 and X_2 are identical to the pre-determined X_1 and X_2, this means that the transitory component in Equation(6.34) is equal to 1 and there is no martingale component in the pricing kernel. To compare the recovered result and true state variables, using the same input parameters a, σ_{X_1} and σ_{X_2} in Borovička et al. (2016), I plot the true joint stationary distribution of the state vector $[X_1, X_2]$ in Figure 6.4.1. It can be seen that the recovered distribution shares similar characteristics with the true distribution. However, the recovered distribution exhibits a lower mean growth rate X_1 and a higher conditional volatility X_2. This result indicates that there exits a martingale component that defines a new probability and under which the distribution of state variables is distorted. The mathematical proof of the new set state variables can be found in Appendix of Borovička et al. (2016).

6.5 Summary

In this chapter, I study the second main problem addressed in the thesis: what is recovered from The Recovery Theorem (TRT). Although previous studies by Audrino et al. (2015) and Jackwerth and Menner (2017) have explored this problem in the equity option market, I find their studies pose two issues: first, they both focus on discrete TRT by Ross (2015), which is sensitive to the number of states and boundaries of state space. Second, the error associated with state price estimation procedure may lead to a misspecified recovery result.

In order to solve these issues, I empirically implement Walden (2017)'s unbounded diffusion Recovery Theorem using S&P 500 index options. In particular, I test whether 1-month realized S&P 500 returns are drawn from recovered probability. My results strongly reject the null hypothesis and confirm Jackwerth and Menner (2017)'s finding using discrete TRT. To further analyze whether the distorted recovery result is due to state price estimation associated error or a theoretical flaw in TRT, I use simulated state prices generated by the Ornstein–Uhlenbeck (OU) process. As expected by Borovička et al. (2016), the recovered probability still differs from the real-world probability. In Section 6.3, I explore theoretical reason that why The Recovery Theorem fails. Also, instead of using a discrete TRT, I provide alternative evidence for the existence of a martingale component in eigenvector/eigenfunction decomposition of the stochastic discount factor using bivariate unbounded recovery.

This study contributes to previous empirical studies of the recovery problem in several aspects. First, to the best of my knowledge, this is the first empirical study of the recovery of real-world probability with an unbounded diffusion process. Without the influence of state space boundaries, I complete Walden (2017)'s unbounded diffusion TRT framework by providing empirical evidence and show that recovery is possible with the unbounded diffusion process. Second, using an extended quadratic SVM, I provide the empirical evidence for Jackwerth and Menner (2017). I show that using a quadratic loss function in state price estimation does not result in well specified recovery results. Third, I contribute to the empirical studies on multivariate TRT. I empirically implement a bivariate recovery, which provides additional unbounded diffusion evidence for Borovička et al. (2016).

Overall, although The Recovery Theorem starting with Ross (2015) (followed by Carr and Yu (2012), Walden (2017) and other researchers) fails to extract the real-world probability, I nevertheless argue it is a remarkable work and leaves many opportunities for further research. Even if at this early stage I can not discard some unrealistic assumptions in TRT, I strongly believe that some future work will solve it.

Chapter 7

Conclusions, Limitations and Future Research

Not everything that counts can be counted, and not everything that can be counted counts.

—Cameron (1963)[55]

7.1 Summary

This thesis focuses on extract forward-looking information from option prices. In particular, I attempt to extract risk neutral and real-world information from option prices. Risk neutral information is related to the theoretical prices of options and real-world information reflects the information associated with the dynamics of the underlying. To extract the risk neutral information, I apply a linear programming support vector machine L_1-SVM framework. Using S&P 500 index options, I show that this L_1-SVM framework is a universal approach and comparatively accurate and smooth. Moreover, using a modified L_1-SVM framework, I further demonstrate how to back out the real-world information from option prices. I empirically investigate The Recovery Theorem (TRT) in an unbounded diffusion setting. Consist with previous empirical study by Jackwerth

[55] This quote is often attributed to Albert Einstein.

and Menner (2017) and theoretical criticism of Borovička et al. (2016), I demonstrate that the recovered probability from TRT is not the real-world probability.

The thesis contains 7 Chapters. Chapter 1 introduces the research background, motivation and key contributions. I briefly summarize the history of options, noting the introduction of electronic trading platforms and creation of more option products as the options markets have become more active and important. Thus, how to obtain and use the information from option markets has been a fundamental concern of both scholars and practitioners. In this thesis, I focus on how to extract accurate information from an option market and the use this information for future research. Specifically, I investigate two problems: (a) how to extract a well-behaved risk neutral density from option prices and (b) whether the information extracted from The Recovery Theorem is the real-world measure.

Chapter 2 provides a brief review of related option pricing and general asset pricing. I illustrate how the risk neutral and real-world measures are related from mathematical finance and asset pricing perspectives. In Section 2.3, I provide a comprehensive review of no arbitrage conditions for call option prices. An introduction to machine learning and its applications in finance is given in Chapter 3, where two machine learning techniques: the neural network and the support vector machine are presented. Comparing their characteristics in terms of ability to handle noisy data, processing large datasets, controlling model complexity, predictive accuracy and ease of operation, I argue the support vector machine is more suitable for extracting the risk neutral density from option prices.

Chapter 4 develops a data filter approach based on three principles: representative, accurate and no arbitrage. In Section 4.2, I provide a detail description of data filter rules and discuss the reason for choosing to work with S&P 500 index options. Comparing my filtered result with Zhang and Xiang (2008), I show that my dataset contains less noise. Finally, to compare my L_1-SVM with other non-parametric methods, I change the estimated framework under the forward measure using the Radon-Nikodym derivative. This transformation enables us to compare all the models on common ground and treat interest rate and dividends as zero.

In light of the above techniques and the theorem, in Chapter 5, I specify how to incorporate no-arbitrage conditions into the linear programming support vector framework. Compared with other methods, I show that previous studies that use different interpolation techniques can all be incorporated in my framework by choosing a different kernel function. Also, taking advantage of the error insensitive range in the support vector machine, my framework naturally considers the information in bid-ask spread. Using 16 years of daily S&P 500 index options, I show that my method establishes a somewhat better accuracy and is stability smooth in the forward-moneyness direction.

In Chapter 6, following Walden (2017), I provide the first empirical study of recovery of real-world probability with the unbounded diffusion process. Modifying my support vector machine framework to consider the quadratic penalty term, I document that using a quadratic loss function in state price estimation does not guarantee a well-recovered result. This distorted result may be caused by either the implementation error or by a theoretical flaw in TRT. To show which is the case, I apply TRT using simulated data from an Ornstein–Uhlenbeck (OU) process. In line with my finding in S&P 500 index options, I note the recovered results always differ from the real-world probability. Following Borovička et al. (2016)'s criticism of Ross (2015), in Section 6.3, I further show the existence of a martingale component in a two-dimensional unbounded diffusion setting.

7.2 Limitations and Future Work

7.2.1 Limitations

Although my study has successfully shown how to extract the risk neutral density and natural probability from option prices, I argue it has some limitations.

First, a data limitation. Although, as shown in Figure 5.4.1, my dataset is large enough to construct a well-defined surface, I argue that high-frequency data are preferred for machine learning approach. This is because, as discussed in Section 5.4.2, the dataset always needs to be split during the learning process.

Second, I fail to address the risk neutral density tail extrapolation problem. As suggested in Section 5.1, I identify five challenges that are imposed in extracting RND. The third issue, that RND should theoretically lie in $[0, +\infty]$, is not appropriately investigated in L_1-SVM framework. In Equation (5.27)-(5.28), I only consider the zero point in my framework because the Python Cvxopt library produces NaN when I use infinity as input. I suggest that future study could apply a different library to solve this issue.

Third, in Chapter 6, while I show empirically that the recovered probability differs from the real-world probability, I argue this does not necessarily mean I should discard the forecasting ability of recovered probability. The proof of the existence of a martingale component using bivariate TRT only statistically implies these two probabilities are not the same. However, it is still possible that partial real-world information is recovered by TRT. How to prove this argument and identify this partial information is an interesting research direction.

7.2.2 Future Research

Clearly, building on Breeden and Litzenberger (1978) and Ross (2015) 's theoretical foundation, various future researches can be explored in order to get option implied information. Based on the content of this thesis, I recommend following research directions.

- It is worth extending the study to more complex options such as American options. This is because, even though the European option is widely traded in the market, in some cases investors prefer to use other types of option to express their future view. For example, the most traded crude oil option traded on the CME is American-style.

- Although the recovered probability is distorted, it still sheds light on applying new extracted probability in risk management and portfolio selection. Given the limited forecast ability of historical data, the estimation of tail risk from past returns has been criticized by researchers. The forward-looking information implied from options, which more closely represents a market's perspective on the future, may yield better performance. I suggest the future research can start with reconstructing the standard risk measure value at the risk (VAR) using implied probability. However, as discussed in Carr and Yu (2012), one issue that should be borne in mind is that the recovered probability reflects the market's view about the future but the market is not always right. Incorrect views obviously would lead incorrect extracted measures.

Bibliography

Aït-Sahalia, Y., Duarte, J., 2003. Nonparametric option pricing under shape restrictions. Journal of Econometrics 116 (1-2), 9–47.

Aït-Sahalia, Y., Lo, A. W., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. Journal of Finance 53 (2), 499–547.

Aït-Sahalia, Y., Lo, A. W., 2000. Nonparametric risk management and implied risk aversion. Journal of Econometrics 94 (1-2), 9–51.

Alvarez, F., Jermann, U. J., 2005. Using asset prices to measure the persistence of the marginal utility of wealth. Econometrica 73 (6), 1977–2016.

Anagnou, I., Bedendo, M., Hodges, S. D., Tompkins, R., 2002. The relation between implied and realised probability density functions. In: Woking Paper,University of Technology,Vienna.

Andersen, M., Dahl, J., Liu, Z., Vandenberghe, L., 2011. Interior-point methods for large-scale cone programming. Optimization for machine learning 5583.

Andricopoulos, A. D., Widdicks, M., Duck, P. W., Newton, D. P., 2003. Universal option valuation using quadrature methods. Journal of Financial Economics 63 (3), 447–471.

Andricopoulos, A. D., Widdicks, M., Newton, D. P., Duck, P. W., 2007. Extending quadrature methods to value multi-asset and complex path dependent options. Journal of Financial Economics 83 (2), 471–499.

Aristotle, 1999. Politics. Batoche Books.

Arrow, K. J., 1964. The role of securities in the optimal allocation of risk-bearing. The Review of Economic Studies 31 (2), 91–96.

Audrino, F., Huitema, R., Ludwig, M., 2015. An empirical analysis of the Ross recovery theorem. Available at SSRN 2433170.

Bakshi, G., Chabi-Yo, F., Gao, X., 2017. A recovery that we can trust? Deducing and testing the restrictions of the recovery theorem. The Review of Financial Studies 31 (2), 532–555.

Bakshi, G., Kapadia, N., Madan, D., 2003. Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options. Review of Financial Studies 16 (1), 101–143.

Bali, T. G., Murray, S., 2013. Does Risk-Neutral Skewness Predict the Cross-Section of Equity Option Portfolio Returns? Journal of Financial and Quantitative Analysis 48 (04), 1145–1171.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. Journal of Finance 59 (4), 1481–1509.

Barberis, N., Thaler, R., 2003. Chapter 18 A survey of behavioral finance. Handbook of the Economics of Finance 1, 1053–1128.

Barletta, A., Santucci de Magistris, P., Sloth, D., 2017. It Only Takes a Few Moments to Hedge. SSRN Electronic Journal.

Bennell, J. A., Sutcliffe, C. M., 2003. Black-Scholes Versus Artificial Neural Networks in Pricing FTSE 100 Options. SSRN Electronic Journal.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business & Economic Statistics 19 (4), 465–474.

Birkhoff, G., de Boor, C., 1965. Piecewise polynomial interpolation and approximation. Approximation of functions, 164–190.

Bishop, C. M., 2006. Pattern recognition and machine learning. springer.

Black, F., 1976. The pricing of commodity contracts. Journal of Financial Economics 3 (1-2), 167–179.

Black, F., 1990. "Living Up To the Model.". Risk Magazine 3, 11–13.

Black, F., Scholes, M., 1973. The Pricing of Options and Corporate Liabilities. Journal of Political Economy 81 (3), 637–654.

Bliss, R. R., Panigirtzoglou, N., 2004. Option-Implied Risk Aversion Estimates. Journal of Finance 59 (1), 407–446.

Borovička, J., Hansen, L. P., Scheinkman, J. A., 2016. Misspecified recovery. The Journal of Finance 71 (6), 2493–2544.

Box, G. E. P., 1976. Science and statistics. Journal of the American Statistical Association 71 (356), 791–799.

Boyle, P. P., 1977. Options: A Monte Carlo approach. Journal of Financial Economics 4 (3), 323–338.

Breeden, D. T., Litzenberger, R. H., 1978. Prices of State-Contingent Claims Implicit in Option Prices. The Journal of Business 51 (4), 621.

Brennan, M. J., Schwartz, E. S., 1977. Finite Difference Methods and Jump Processes Arising in the Pricing of Contingent Claims: A Synthesis. Journal of Financial and Quantitative Analysis 13 (3), 461–474.

Buchen, P. W., Kelly, M., 1996. The Maximum Entropy Distribution of an Asset Inferred from Option Prices. The Journal of Financial and Quantitative Analysis 31 (1), 143.

Cameron, W. B., 1963. Informal sociology: A casual introduction to sociological thinking. Vol. 21. Random House.

Carr, P., Madan, D. B., 2005. A note on sufficient conditions for no arbitrage. Finance Research Letters 2 (3), 125–130.

Carr, P., Yu, J., 2012. Risk, Return, and Ross Recovery. The Journal of Derivatives 20, 38–59.

Chen, D., Härkönen, H. J., Newton, D. P., 2014. Advancing the universality of quadrature methods to any underlying process for option pricing. Journal of Financial Economics 114 (3), 600–612.

Chen, M.-Y., 2014. Using a hybrid evolution approach to forecast financial failures for Taiwan-listed companies. Quantitative Finance 14 (6), 1047–1058.

Chen, S., Härdle, W. K., Moro, R. A., 2011. Modeling default risk with support vector machines. Quantitative Finance 11 (1), 135–154.

Chiang, M.-H., Chiu, H.-Y., Chou, R. K., 2016. Measuring the disposition effect on the option market: New evidence. Working Paper,National Chengchi University.

Christoffersen, P., Heston, S., Jacobs, K., 2013. Capturing option anomalies with a variance-dependent pricing kernel. Review of Financial Studies 26 (8), 1962–2006.

Constantinides, G. M., Jackwerth, J. C., Savov, A., 2013. The Puzzle of Index Option Returns. Review of Asset Pricing Studies 3 (2), 229–257.

Cont, R., 1998. Beyond implied volatility. Econophysics, 26.

Corrado, C. J., 2001. Option pricing based on the generalized lambda distribution. Journal of Futures Markets 21 (3), 213–236.

Cox, J. C., Ross, S. A., Rubinstein, M., 1979. Option pricing: A simplified approach. Journal of Financial Economics 7 (3), 229–263.

Cui, L., Huang, K., Cai, H. J., 2015. Application of a TGARCH-wavelet neural network to arbitrage trading in the metal futures market in China. Quantitative Finance 15 (2), 371–384.

Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. Mathematische Annalen 300 (1), 463–520.

Delbaen, F., Schachermayer, W., 1998. The fundamental theorem of asset pricing for unbounded stochastic processes. Mathematische Annalen 312, 215–250.

DeMiguel, V., Plyakha, Y., Uppal, R., Vilkov, G., 2013. Improving Portfolio Selection Using Option-Implied Volatility and Skewness. Journal of Financial and Quantitative Analysis 48 (6), 1813–1845.

Ding, S., Li, H., Su, C., Yu, J., Jin, F., 2013. Evolutionary artificial neural networks: a review. Artificial Intelligence Review 39 (3), 251–260.

Donaldson, R. G., Kamstra, M., 1997. An artificial neural network-GARCH model for international stock return volatility. Journal of Empirical Finance 4 (1), 17–46.

Dubynskiy, S., Goldstein, R. S., 2013. Recovering Drifts and Preference Parameters from Financial Derivatives. SSRN Electronic Journal.

Dybvig, P. H., Ross, S. A., 1989. Arbitrage. In: Finance. Springer, pp. 57–71.

Elad, M. M., 2010. Sparse and redundant representations : from theory to applications in signal and image processing. Springer.

Engle, R. F., Granger, C. W. J., 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. Econometrica, 251–276.

Fabozzi, F. J., Tunaru, R., Albota, G., 2009. Estimating risk-neutral density with parametric models in interest rate markets. Quantitative Finance 9 (1), 55–70.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance 25 (2), 383–417.

Fengler, M. R., 2009. Arbitrage-free smoothing of the implied volatility surface. Quantitative Finance 9 (4), 417–428.

Fengler, M. R., Hin, L. Y., 2015. Semi-nonparametric estimation of the call-option price surface under strike and time-to-expiry no-arbitrage constraints. Journal of Econometrics 184 (2), 242–261.

Figlewski, S., 2008. Estimating the Implied Risk-Neutral Density for the US Market Portfolio. In: Volatility and Time Series Econometrics: Essays in Honor of Robert Engle. Oxford University Press, Oxford,UK, pp. 323–354.

Figlewski, S., 2018. Risk Neutral Densities: A Review. SSRN Electronic Journal.

Giacomini, R., Gottschling, A., Haefke, C., White, H., 2008. Mixtures of t-distributions for finance and forecasting. Journal of Econometrics 144 (1), 175–192.

Glaser, J., Heider, P., 2012. Arbitrage-free approximation of call price surfaces and input data risk. Quantitative Finance 12 (1), 61–73.

Goodfellow, I., Bengio, Y., Courville, A., Goodfellow, Ian, Bengio, Yoshua, Courville, A., 2016. Deep Learning. MIT Press.

Gope, P., Fries, C. P., 2011. Arbitrage-free Asset Class Independent Volatility Surface Interpolation on Probability Space using Normed Call Prices. SSRN Electronic Journal, 1–27.

Granger, C. W., 2005. The past and future of empirical finance: some personal comments. Journal of Econometrics 129 (1-2), 35–40.

Green, P. J., Silverman, B. W., 1994. Nonparametric regression and generalized linear models. Number 58 in Monographs on Statistics and Applied Probability. No. 58.

Hammer, B., Gersmann, K., 2003. A note on the universal approximation capability of support vector machines. Neural Processing Letters 17 (1), 43–53.

Hansen, L. P., 2013. Uncertainty outside and inside economic models. Nobel Lecture: https://www. nobelprize. org/nobel_prizes/economicsciences/laureates/2013/hansen-lecture. pdf 122 (5).

Hansen, L. P., Heaton, J., Lee, J., Roussanov, N., 2007. Intertemporal substitution and risk aversion. Handbook of econometrics 6, 3967–4056.

Hansen, L. P., Scheinkman, J., 2009. Long-Term Risk: An Operator Approach. Econometrica 77 (1), 177–234.

Hentschel, L., 2003. Errors in Implied Volatility Estimation. The Journal of Financial and Quantitative Analysis 38 (4), 779–810.

Heston, S. L., 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. Review of Financial Studies 6 (2), 327–343.

Hinton, G. E., Osindero, S., Teh, Y.-W., 2006. A Fast Learning Algorithm for Deep Belief Nets. Neural Computation 18 (7), 1527–1554.

Huang, D., Shaliastovich, I., 2014. Risk adjustment and the temporal resolution of uncertainty: evidence from options markets. Working Paper, University of Pennsylvania.

Hull, J. C., Basu, S., 2016. Options, futures, and other derivatives. Pearson Education India.

Hutchinson, J. M., Lo, A. W., Poggio, T., 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. The Journal of Finance 49 (3), 851–889.

Jackwerth, J. C., 1999. Option-Implied Risk-Neutral Distributions and Implied Binomial Trees. The Journal of Derivatives 7 (1999), 66–82.

Jackwerth, J. C., 2000. Recovering risk aversion from option prices and realized returns. Review of Financial Studies 13 (2), 433–451.

Jackwerth, J. C., Rubinstein, M., 1996. Recovering Probability Distributions from Option Prices. The Journal of Finance 51 (5), 1611–1631.

Jackwerth, J. C. J., Menner, M., 2017. Does the Ross Recovery Theorem Work Empirically? Working Paper, University of Konstanz.

Jensen, J. L. W. V., 1906. Sur les fonctions convexes et les inegalites entre les valeurs moyennes. Acta Mathematica 30 (1), 175–193.

Jiang, G. J., Tian, Y. S., 2005. The model-free implied volatility and its information content. Review of Financial Studies 18 (4), 1305–1342.

Jondeau, E., Rockinger, M., 2000. Reading the smile: the message conveyed by methods which infer risk neutral densities. Journal of International Money and Finance 19 (6), 885–915.

Jondeau, E., Rockinger, M., 2001. Gram-Charlier densities. Journal of Economic Dynamics and Control 25 (10), 1457–1483.

Kelly, J. J. L., 1956. A New Interpretation of Information Rate. Bell System Technical Journal 35 (4), 917–926.

Knight, F. H., 1921. Risk, uncertainty and profit. University of Chicago Press, Chicago.

Konishi, S., Kitagawa, G., 2008. Information criteria and statistical modeling. Springer Science & Business Media.

Kreps, D. M., Porteus, E. L., 1978. Temporal resolution of uncertainty and dynamic choice theory. Econometrica 46 (1), 185–200.

Kundu, A., Kumar, S., Kumar Tomar, N., Kumar Gupta, S., 2016. Call option price function in Bernstein polynomial basis with no-arbitrage inequality constraints. Journal of Inequalities and Applications 2016 (1), 153.

Lai, W.-N., 2011. Comparison of methods to estimate option implied risk-neutral densities. Quantitative Finance (December 2014), 1–17.

Liang, X., Zhang, H., Xiao, J., Chen, Y., 2009. Improving option price forecasts with neural networks and support vector regressions. Neurocomputing 72 (13-15), 3055–3065.

Lim, G. C., Martin, G. M., Martin, V. L., 2005. Parametric pricing of higher order moments in S&P500 options. Journal of Applied Econometrics 20 (3), 377–404.

Liu, F., 2014. Recovering Conditional Return Distributions by Regression: Estimation and Applications. Available at SSRN 2530183.

Long, J., 1990. The numeraire portfolio. Journal of Financial Economics 26 (1), 29–69.

Lucas, R. E. J., 1978. Asset Prices in an Exchange Economy. Econometrica, 1429–1445.

Ludwig, M., 2015. Robust Estimation of Shape Constrained State Price Density Surfaces. Journal of Derivatives 22 (3), 56–72.

Luss, R., D'Aspremont, A., 2015. Predicting abnormal returns from news using text classification. Quantitative Finance 15 (6), 999–1012.

Marsh, A., 2018. Citi Seeks Credit Markets Edge With New Fintech Investment Group.

URL https://www.bloomberg.com/news/articles/2018-09-05/citi-s-credit-unit-starts-fintech-investment-group-under-zhang

McCulloch, W. S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics 5 (4), 115–133.

Mehra, R., Prescott, E. C., 1985. The equity premium: A puzzle. Journal of monetary Economics 15 (2), 145–161.

Melick, W. R., Thomas, C. P., 1997. Recovering an Asset's Implied PDF from Option Prices: An Application to Crude Oil during the Gulf Crisis. The Journal of Financial and Quantitative Analysis 32 (1), 91–115.

Mercer, J., 1909. Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 209, 441–458.

Merton, R. C., 1973. Theory of Rational Option Pricing. The Bell Journal of Economics and Management Science 4 (1), 141.

Meyer, C., 2000. Matrix analysis and applied linear algebra. Society for Industrial and Applied Mathematics.

Minsky, M., Papert, S., 1960. Perceptrons: expanded edition. MIT Press Cambridge MA.

Monnier, J.-B., 2013. Risk-neutral density recovery via spectral analysis. SIAM Journal on Financial Mathematics 4 (1), 650–667.

Myron Scholes, 2016. Black-Scholes and beyond. Interview with Myron Scholes - YouTube.
URL https://www.youtube.com/watch?v=2Axx5f2WgHc

Pang, B., Lee, L., Vaithyanathan, S., 2002. Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, pp. 79–86.

Park, H., 2016. Ross recovery with recurrent and transient processes. Quantitative Finance 16 (5), 667–676.

Patterson, J., Gibson, A., 2017. Deep Learning: A Practitioner's Approach. O'Reilly Media.

Pérez-cruz, F., Afonso-rodríguez, J. A., Giner, J., 2003. Estimating GARCH models using support vector machines. Quantitative Finance 3 (3), 163–172.

Platen, E., 2006. A Benchmark approach to finance. Mathematical Finance 16 (1), 131–151.

Polkovnichenko, V., Zhao, F., 2013. Probability weighting functions implied in options prices. Journal of Financial Economics 107 (3), 580–609.

Poon, S.-H., Rockinger, M., Tawn, J., 2003. Extreme value dependence in financial markets: Diagnostics, models, and financial implications. The Review of Financial Studies 17 (2), 581–610.

Qi, M., Maddala, G. S., 1996. Option pricing using artificial neural networks: The case of S&P 500 index call options. In: InNeural networks in financial engineering: proceedings of the Third International Conference on Neural Networks in the Capital Markets. New York: World Scientific., pp. 78–91.

Qin, L., Linetsky, V., 2016. Positive Eigenfunctions of Markovian Pricing Operators: Hansen-Scheinkman Factorization, Ross Recovery, and Long-Term Pricing. Operations Research 64 (1), 99–117.

Qin, L., Linetsky, V., 2017. Long-Term Risk: A Martingale Approach. Econometrica 85 (1), 299–312.

Qin, L., Linetsky, V., Nie, Y., jan 2016. Long Forward Probabilities, Recovery and the Term Structure of Bond Risk Premiums. SSRN Electronic Journal.

Rockinger, M., Jondeau, E., 2002. Entropy densities with an application to autoregressive conditional skewness and kurtosis. Journal of Econometrics 106 (1), 119–142.

Rompolis, L. S., Tzavalis, E., 2009. Recovering Risk Neutral Densities from Option Prices: A New Approach. Journal of Financial and Quantitative Analysis 43 (4), 1037–1053.

Roper, M., 2010. Arbitrage Free Implied Volatility Surfaces. Working paper, School of Mathematics and Statistics, University of Sydney, NSW, Australia.

Rosenberg, J. V., 1998. Pricing multivariate contingent claims using estimated risk-neutral density functions. Journal of International Money and Finance 17 (2), 229–247.

Rosenberg, J. V., Engle, R. F., 2002. Empirical pricing kernels. Journal of Financial Economics 64 (3), 341–372.

Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review 65 (6), 386.

Ross, S., 2015. The Recovery Theorem. Journal of Finance 70 (2), 1–61.

Ross, S. A., 2009. Neoclassical finance. Princeton University Press.

Rubinstein, M., 1985. Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. The Journal of Finance 40 (2), 455–480.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533.

Sanford, A., 2017. Recovery theorem with a multivariate markov chain. Tech. rep.

Schölkopf, B., Smola, A. J., 2002. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

Stefanek, G., 1987. The handbook of artificial intelligence. Computers & Mathematics with Applications.

Tam, K. Y., Kiang, M. Y., 1992. Managerial applications of neural networks: the case of bank failure predictions. Management science 38 (7), 926–947.

Thompson, E. A., 2007. The tulipmania: Fact or artifact? Public Choice 130 (1-2), 99–114.

Tran, N.-K., Xia, S., 2015. Specified recovery. Unpublished working paper. Washington University in St. Louis.

Vapnik, V., Golowich, S. E., Smola, A., 1997. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. In: Advances in Neural Information Processing Systems 9. pp. 281–287.

Vapnik, V., Lerner, A., 1963. Pattern recognition using generalized portrait method. Automation and remote control.

Vapnik, V. N., 2000. The Nature of Statistical Learning Theory. Springer New York, New York, NY.

Vapnik, V. N., Chervonenkis, A. Y., 2015. On the uniform convergence of relative frequencies of events to their probabilities. In: Measures of Complexity. Springer,Cham. pp. 11–30.

Walden, J., 2017. Recovery with unbounded diffusion processes. Review of Finance 21 (4), 1403–1444.

Werbos, P. J., 1974. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.

Wilmott, P., 2006. Paul wilmott on quantitative finance. Wiley.

Yatchew, A., Härdle, W., 2006. Nonparametric state price density estimation using constrained least squares and the bootstrap. Journal of Econometrics 133 (2), 579–599.

Zhang, J. E., Xiang, Y., 2008. The implied volatility smirk. Quantitative Finance 8 (3), 263–284.

Appendix

Kernel Function

Let X and Y are input data, the parameter γ determines the amplitude on the function. The popular choice of kernel function $K(X,Y)$ are listed as following:

- Polynomial kernel with degree c:

$$K(X,Y) = (\gamma X^T Y + \gamma)^c$$

- Linear kernel:

$$K(X,Y) = X^T Y$$

- Radial basis function(RBF) kernel:

$$K(X,Y) = exp(||X - Y||^2/2\sigma^2)$$

- Sigmoid kernel:

$$K(X,Y) = tanh(\gamma.X^T Y + r)$$

Selected Python Code

Selected Code for Data Cleaning

```
import numpy as np
import pandas as pd
import datetime
import matplotlib.pyplot as plt
import dateutil
import re
from pandas import DataFrame
from scipy.interpolate import interp1d
from scipy.stats import norm
from scipy.linalg import lstsq

# 1.import option data
df_option = pd.read_csv('/Users/Desktop/original data copy/
    spx2000-20160430.csv')
# Note that the strike prices are out by a factor of 1,000!  So
    let's fix them: df_option.strike_price = df_option.
    strike_price/1000
# format date
df_option['date'] = pd.to_datetime(df_option['date'],format = '%m
    /%d/%Y')
df_option['exdate'] = pd.to_datetime(df_option['exdate'],format =
     '%m/%d/%Y')
df_option['days']=(df_option['exdate']-df_option['date']).dt.days

# 2.import index data
df_index = pd.read_csv('/Users/chelsea/Desktop/all thesis/
    original data copy/chapter1data/spxunderlying2000-2016.csv')
df_index['date'] = pd.to_datetime(df_index['date'],format = '%m/%
    d/%Y') df_index = df_index.rename(columns = {'close':'
```

```
    index_price','volume':'underlying_volume'})

# final dataset
df = pd.merge(df_option,df_index)
df['moneyness']=df['strike_price']/df['index_price'] df['
    mid_price']=0.5*(df['best_bid']+df['best_offer'])

# drop identical observation
print(len(df.drop_duplicates(subset=['date','exdate','
    strike_price','cp_flag','best_bid','best_offer'],keep='first')
    ))

# drop Identical except price
df_i = df.drop_duplicates(subset=['date','exdate','strike_price',
    'cp_flag'],keep='first') len(df_i)

# Liquidity Filter
df_zero = df_i[(df_i['best_bid']!=0) & (df_i['best_offer']!=0) &
    (df_i['volume']!=0)]
df_days = df_zero[(df_zero['days']>= 7) & (df_zero['days']<= 365)
    ]
df_f = df_days[(df_days['moneyness']>= 0.8) & (df_days['moneyness
    ']<= 1.2)]
```

Selected Code for L_1-SVM

```
import numpy as np
import pandas as pd
import sympy as sp
from sympy import exp
from sympy.utilities.lambdify import lambdify
from cvxopt import matrix, solvers
solvers.options['show_progress'] = False
from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV, cross_val_score, KFold
from sklearn.base import BaseEstimator, RegressorMixin
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import make_scorer

# kernels
def linear_ker(x, y):
    return np.dot(x, y)

def poly_ker(x, y, p=3):
    return (1 + np.dot(x, y)) ** p

def rbf_ker(x,y,sigma):
    return exp(-np.dot(x-y,x-y)/ (2 * (sigma ** 2)))

def sigmoid_ker(x, y, p1=1,p2=2):
    return sp.tanh(p1*np.dot(x, y) + p2)

## differential function
def diff_fun(zinput, k, ker, xj,vars):
    '''
    k is order
```

```
    ker:kernel function choice
    zinput:knots in spline
    xj:jth compoent of X
    vars: sigma if rbf_kernel
    '''
    sizeX = len(xj)
    sizezinput = len(zinput)
    sx = sp.symbols('x0:%d' % sizeX)
    x0 = sp.symbols('x0')  # change to x1, whether results change???
    difftmp = np.empty([sizezinput, sizeX], dtype=object)

    for i in np.arange(sizezinput):
        for j in np.arange(sizeX):
            difftmp[i][j] = ker(zinput[i], sx[j],vars)

    diff_fun = sp.diff(sp.Matrix(difftmp), x0, k)
    dffun = diff_fun.col(0)
    yf = lambdify(x0, dffun, 'numpy')
    res = yf(xj)
    n = res.shape[0]
    taok = res.reshape(n, -1)

    return taok

# Optimazation

class mySVR(BaseEstimator, RegressorMixin):
    '''
    Support vector regression and No arbitarge constraints
    '''
    def __init__(self, e=0.25, ker=rbf_ker, nsplit=2, staus='Non_Cons', C=100,
```

```
                 lam0=50,lam1=50,lam2=50,sigma=0.707,tol = 1e-10):
        self.e = e
        self.ker = ker
        self.nsplit = nsplit
        self.staus = staus
        self.C = C
        self.lam0 = lam0
        self.lam1 = lam1
        self.lam2 = lam2
        self.sigma = sigma
        self.tol = tol

    def fit(self, X, y):
        '''
        my SVR eastimator
        y-output vector
        x-input vector
        ker is kernel, choices:[linear_ker, poly_ker, rbf_ker,
                        sigmoid_ker, tensorproduct, b-spline]
        C, e, lam0 are trade-off parameters
        set e 1/4(bid-ask spread)
        C=1, tol =1e-10 tolerance for support vector detection
        nsplit-no. of splited fwd points
        staus: whether add NA constraions or not. Flase is Non_Cons, Ture = Cons
        '''
        n = len(X)
        K = kernelmatrix(X, self.ker, self.sigma)
        y = y.reshape(len(y), -1)

        # define the linear programing
        if self.staus == 'Non_Cons':
```

```
c = matrix(np.vstack((np.zeros((n, 1)), 0, (1/n)*self.C
        * np.ones((n, 1)), (1/n)*np.ones((n, 1)))),tc='d')
```

S&P 500 Index Sample List

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
1	ABT	ABBOTT LABORATORIES	31-Mar-1964		19023	2834	Pharmaceutical Preparations
2	HON	HONEYWELL INTERNATIONAL INC	31-Mar-1964		19023	9997	
3	AEP	AMERICAN ELECTRIC POWER CO	31-Mar-1964		19023	4911	Electric Services
4	BA	BOEING CO	31-Mar-1964		19023	3721	Aircraft
5	BMY	BRISTOL-MYERS SQUIBB CO	31-Mar-1964		19023	2834	Pharmaceutical Preparations
6	CPB	CAMPBELL SOUP CO	31-Mar-1964		19023	2030	Canned, Frozen & Preserved Fruit, Veg & Food Specialties
7	CAT	CATERPILLAR INC	31-Mar-1964		19023	3531	Construction Machinery & Equip
8	CVX	CHEVRON CORP	31-Mar-1964		19023	2911	Petroleum Refining
9	KO	COCA-COLA CO	31-Mar-1964		19023	2086	Bottled & Canned Soft Drinks & Carbonated Waters
10	CL	COLGATE-PALMOLIVE CO	31-Mar-1964		19023	2844	Perfumes, Cosmetics & Other Toilet Preparations
11	ED	CONSOLIDATED EDISON INC	31-Mar-1964		19023	4931	Electric & Other Services Combined
12	GLW	CORNING INC	31-Mar-1964		19023	3679	Electronic Components, NEC
13	DE	DEERE & CO	31-Mar-1964		19023	3523	Farm Machinery & Equipment

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
14	DTE	DTE ENERGY CO	31-Mar-1964		19023	4931	Electric & Other Services Combined
15	D	DOMINION ENERGY INC	31-Mar-1964		19023	4911	Electric Services
16	DWDP	DOWDUPONT INC	31-Mar-1964		19023	2821	Plastic Materials, Synth Resins & Nonvulcan Elastomers
17	ETN	EATON CORP PLC	31-Mar-1964		19023	3620	Electrical Industrial Apparatus
18	XOM	EXXON MOBIL CORP	31-Mar-1964		19023	2911	Petroleum Refining
19	F	FORD MOTOR CO	31-Mar-1964		19023	3711	Motor Vehicles & Passenger Car Bodies
20	GD	GENERAL DYNAMICS CORP	31-Mar-1964		19023	3721	Aircraft
21	GE	GENERAL ELECTRIC CO	31-Mar-1964		19023	9997	
22	GT	GOODYEAR TIRE & RUBBER CO	31-Mar-1964		19023	3011	Tires & Inner Tubes
23	HAL	HALLIBURTON CO	31-Mar-1964		19023	1389	Oil & Gas Field Services, NEC
24	HSY	HERSHEY CO	31-Mar-1964		19023	2060	Sugar & Confectionery Products
25	IBM	INTL BUSINESS MACHINES CORP	31-Mar-1964		19023	7370	Services-Computer Programming, Data Processing, Etc.
26	IP	INTL PAPER CO	31-Mar-1964		19023	2631	Paperboard Mills
27	K	KELLOGG CO	31-Mar-1964		19023	2040	Grain Mill Products
28	KMB	KIMBERLY-CLARK CORP	31-Mar-1964		19023	2621	Paper Mills
29	KR	KROGER CO	31-Mar-1964		19023	5411	Retail-Grocery Stores
30	SPGI	S&P GLOBAL INC	31-Mar-1964		19023	7323	

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
31	CVS	CVS HEALTH CORP	31-Mar-1964		19023	5912	Retail-Drug Stores and Proprietary Stores
32	MRK	MERCK & CO	31-Mar-1964		19023	2834	Pharmaceutical Preparations
33	ETR	ENTERGY CORP	31-Mar-1964		19023	4911	Electric Services
34	MMM	3M CO	31-Mar-1964		19023	2670	Converted Paper & Paperboard Prods (No Containers/Boxes)
35	MSI	MOTOROLA SOLUTIONS INC	31-Mar-1964		19023	3663	Radio & TV Broadcasting & Communications Equipment
36	NSC	NORFOLK SOUTHERN CORP	31-Mar-1964		19023	4011	Railroads, Line-Haul Operating
37	XEL	XCEL ENERGY INC	31-Mar-1964		19023	4931	Electric & Other Services Combined
38	FE	FIRSTENERGY CORP	31-Mar-1964		19023	4911	Electric Services
39	PPG	PPG INDUSTRIES INC	31-Mar-1964		19023	2851	Paints, Varnishes, Lacquers, Enamels & Allied Prods
40	PCG	PG&E CORP	31-Mar-1964		19023	4931	Electric & Other Services Combined
41	SRE	SEMPRA ENERGY	31-Mar-1964		19023	4931	Electric & Other Services Combined
42	PEP	PEPSICO INC	31-Mar-1964		19023	2080	Beverages
43	PFE	PFIZER INC	31-Mar-1964		19023	2834	Pharmaceutical Preparations
44	EXC	EXELON CORP	31-Mar-1964		19023	4911	Electric Services
45	MO	ALTRIA GROUP INC	31-Mar-1964		19023	2111	Cigarettes

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
46	COP	CONOCOPHILLIPS	31-Mar-1964		19023	1311	Crude Petroleum & Natural Gas
47	PG	PROCTER & GAMBLE CO	31-Mar-1964		19023	2840	Soap, Detergents, Cleaning Preparations, Perfumes, Cosmetics
48	PEG	PUBLIC SERVICE ENTRP GRP INC	31-Mar-1964		19023	4931	Electric & Other Services Combined
49	RTN	RAYTHEON CO	31-Mar-1964		19023	3812	Search, Detection, Navigation, Guidance, Aeronautical Sys
50	ROK	ROCKWELL AUTOMATION	31-Mar-1964		19023	3620	Electrical Industrial Apparatus
51	EIX	EDISON INTERNATIONAL	31-Mar-1964		19023	4911	Electric Services
52	SO	SOUTHERN CO	31-Mar-1964		19023	4911	Electric Services
53	TXN	TEXAS INSTRUMENTS INC	31-Mar-1964		19023	3674	Semiconductors & Related Devices
54	UNP	UNION PACIFIC CORP	31-Mar-1964		19023	4011	Railroads, Line-Haul Operating
55	UTX	UNITED TECHNOLOGIES CORP	31-Mar-1964		19023	3724	Aircraft Engines & Engine Parts
56	WY	WEYERHAEUSER CO	31-Mar-1964		19023	2400	Lumber & Wood Products (No Furniture)
57	WHR	WHIRLPOOL CORP	31-Mar-1964		19023	3630	Household Appliances
58	XRX	XEROX CORP	31-Mar-1964		19023	3577	Computer Peripheral Equipment, NEC
59	D2	CONSOLIDATED NATURAL GAS CO	31-Mar-1964	30-Jan-2000	19023	4923	Natural Gas Transmission & Distribution

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
60	FWLT	FOSTER WHEELER AG	31-Mar-1964	30-Jan-2000	19023	1600	Heavy Construction Other Than Bldg Const - Contractors
61	ARC.3	ATLANTIC RICHFIELD CO	31-Mar-1964	17-Apr-2000	19023	2911	Petroleum Refining
62	RLM.1	REYNOLDS METALS CO	31-Mar-1964	4-May-2000	19023	3334	Primary Production of Aluminum
63	CBS.2	CBS CORP -OLD	31-Mar-1964	4-May-2000	19023	4833	Television Broadcasting Stations
64	NC	NACCO INDUSTRIES -CL A	31-Mar-1964	4-Jun-2000	19023	1221	Bituminous Coal & Lignite Surface Mining
65	CSR.1	CENTRAL & SOUTH WEST CORP	31-Mar-1964	15-Jun-2000	19023	4911	Electric Services
66	WLA	WARNER-LAMBERT CO	31-Mar-1964	20-Jun-2000	19023	2834	Pharmaceutical Preparations
67	MZIAQ	MILACRON INC	31-Mar-1964	28-Jun-2000	19023	3559	Special Industry Machinery, NEC
68	GTE.1	GTE CORP	31-Mar-1964	2-Jul-2000	19023	4813	Telephone Communications (No Radiotelephone)
69	GAPTQ	GREAT ATLANTIC & PAC TEA CO	31-Mar-1964	29-Aug-2000	19023	5411	Retail-Grocery Stores
70	OC	OWENS CORNING	31-Mar-1964	1-Oct-2000	19023	3290	Abrasive, Asbestos & Misc Nonmetallic Mineral Prods
71	BFO.2	BESTFOODS	31-Mar-1964	2-Oct-2000	19023	2030	Canned, Frozen & Preserved Fruit, Veg & Food Specialties
72	UCM.2	UNICOM CORP	31-Mar-1964	22-Oct-2000	19023	4911	Electric Services
73	NI2	COLUMBIA ENERGY GROUP	31-Mar-1964	1-Nov-2000	19023	4923	Natural Gas Transmission & Distribution
74	3ACKH	ARMSTRONG HOLDINGS INC	31-Mar-1964	16-Nov-2000	19023	3089	Plastics Products, NEC

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
75	3BHMSQ	BETHLEHEM STEEL CORP	31-Mar-1964	10-Dec-2000	19023	3312	Steel Works, Blast Furnaces & Rolling Mills (Coke Ovens)
76	CCK	CROWN HOLDINGS INC	31-Mar-1964	10-Dec-2000	19023	3411	Metal Cans
77	GRA	GRACE (W R) & CO	31-Mar-1964	10-Dec-2000	19023	2810	Industrial Inorganic Chemicals
78	VO.3	SEAGRAM CO LTD	31-Mar-1964	10-Dec-2000	19023	3652	Phonograph Records & Prerecorded Audio Tapes & Disks
79	PRDCQ	PRIMARY PDC INC	31-Mar-1964	11-Dec-2000	19023	3861	Photographic Equipment & Supplies
80	DOW2	UNION CARBIDE CORP	31-Mar-1964	6-Feb-2001	19023	2860	Industrial Organic Chemicals
81	OAT.	QUAKER OATS CO	31-Mar-1964	2-Aug-2001	19023	2040	Grain Mill Products
82	TKR	TIMKEN CO	31-Mar-1964	6-Aug-2001	19023	3562	Ball & Roller Bearings
83	OKE	ONEOK INC	31-Mar-1964	29-Aug-2001	19023	4923	Natural Gas Transmission & Distribution
84	TX.2	TEXACO INC	31-Mar-1964	9-Oct-2001	19023	2911	Petroleum Refining
85	ENRNQ	ENRON CORP	31-Mar-1964	29-Nov-2001	19023	5172	Wholesale-Petroleum & Petroleum Products (No Bulk Stations)
86	HM	HOMESTAKE MINING	31-Mar-1964	16-Dec-2001	19023	1040	Gold and Silver Ores
87	FMC	FMC CORP	31-Mar-1964	1-Jan-2002	19023	2870	Agricultural Chemicals
88	SHLD	SEARS HOLDINGS CORP	31-Mar-1964	16-Jan-2002	19023	5311	Retail-Department Stores
89	MEA.1	MEAD CORP	31-Mar-1964	29-Jan-2002	19023	2621	Paper Mills

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
90	NMK	NIAGARA MOHAWK HOLDINGS INC	31-Mar-1964	31-Jan-2002	19023	4931	Electric & Other Services Combined
91	AL.1	ALCAN INC	31-Mar-1964	21-Jul-2002	19023	3350	Rolling Drawing & Extruding of Nonferrous Metals
92	INCLF	INCO LTD	31-Mar-1964	21-Jul-2002	19023	3330	Primary Smelting & Refining of Nonferrous Metals
93	RDPL	ROYAL DUTCH PETROLEUM NV	31-Mar-1964	21-Jul-2002	19023	2911	Petroleum Refining
94	UN	UNILEVER NV	31-Mar-1964	21-Jul-2002	19023	2840	Soap, Detergents, Cleaning Preparations, Perfumes, Cosmetics
95	TRW.1	TRW INC	31-Mar-1964	11-Dec-2002	19023	3714	Motor Vehicle Parts & Accessories
96	AAL	AMERICAN AIRLINES GROUP INC	31-Mar-1964	13-Mar-2003	19023	4512	Air Transportation, Scheduled
97	HBC1	HSBC FINANCE CORP	31-Mar-1964	30-Mar-2003	19023	6141	Personal Credit Institutions
98	PHA.1	PHARMACIA CORP	31-Mar-1964	15-Apr-2003	19023	2834	Pharmaceutical Preparations
99	WINN	WINN-DIXIE STORES INC	31-Mar-1964	2-Dec-2004	19023	5411	Retail-Grocery Stores
100	CR	CRANE CO	31-Mar-1964	19-Dec-2004	19023	3490	Miscellaneous Fabricated Metal Products
101	S.1	SEARS ROEBUCK & CO	31-Mar-1964	27-Mar-2005	19023	5311	Retail-Department Stores
102	UCL	UNOCAL CORP	31-Mar-1964	10-Aug-2005	19023	1311	Crude Petroleum & Natural Gas
103	MAY.2	MAY DEPARTMENT STORES CO	31-Mar-1964	28-Aug-2005	19023	5311	Retail-Department Stores

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
104	G.1	GILLETTE CO	31-Mar-1964	2-Oct-2005	19023	3420	Cutlery, Handtools & General Hardware
105	T.2	AT&T CORP	31-Mar-1964	20-Nov-2005	19023	4813	Telephone Communications (No Radiotelephone)
106	GP.1	GEORGIA-PACIFIC CORP	31-Mar-1964	19-Dec-2005	19023	2600	Papers & Allied Products
107	DAN	DANA INC	31-Mar-1964	2-Mar-2006	19023	3714	Motor Vehicle Parts & Accessories
108	MYG.1	MAYTAG CORP	31-Mar-1964	2-Apr-2006	19023	3630	Household Appliances
109	NAV	NAVISTAR INTERNATIONAL CORP	31-Mar-1964	19-Dec-2006	19023	3711	Motor Vehicles & Passenger Car Bodies
110	PGL.1	PEOPLES ENERGY CORP	31-Mar-1964	21-Feb-2007	19023	4924	Natural Gas Distribution
111	PD	PHELPS DODGE CORP	31-Mar-1964	19-Mar-2007	19023	3330	Primary Smelting & Refining of Nonferrous Metals
112	0033A	ENERGY FUTURE HOLDINGS CORP	31-Mar-1964	9-Oct-2007	19023	4911	Electric Services
113	WWY	WRIGLEY (WM) JR CO	31-Mar-1964	6-Oct-2008	19023	2060	Sugar & Confectionery Products
114	UIS	UNISYS CORP	31-Mar-1964	10-Nov-2008	19023	7373	Services-Computer Integrated Systems Design
115	HPC	HERCULES INC	31-Mar-1964	13-Nov-2008	19023	2890	Miscellaneous Chemical Products
116	GM	GENERAL MOTORS CO	31-Mar-1964	2-Jun-2009	19023	3711	Motor Vehicles & Passenger Car Bodies

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
117	IR	INGERSOLL-RAND PLC	31-Mar-1964	30-Jun-2009	19023	3585	Air-Cond & Warm Air Heatg Equip & Comm & Indl Refrig Equip
118	CBE	COOPER INDUSTRIES PLC	31-Mar-1964	8-Sep-2009	19023	3640	Electric Lighting & Wiring Equipment
119	WYE	WYETH	31-Mar-1964	15-Oct-2009	19023	2834	Pharmaceutical Preparations
120	SGP	SCHERING-PLOUGH	31-Mar-1964	3-Nov-2009	19023	2834	Pharmaceutical Preparations
121	BRK3	BURLINGTON NORTHERN SANTA FE	31-Mar-1964	15-Feb-2010	19023	4011	Railroads, Line-Haul Operating
122	KODK	EASTMAN KODAK CO	31-Mar-1964	19-Dec-2010	19023	3577	Computer Peripheral Equipment, NEC
123	ITT	ITT INC	31-Mar-1964	31-Oct-2011	19023	3561	Pumps & Pumping Equipment
124	CEG	CONSTELLATION ENERGY GRP INC	31-Mar-1964	13-Mar-2012	19023	4931	Electric & Other Services Combined
125	GR	GOODRICH CORP	31-Mar-1964	30-Jul-2012	19023	3728	Aircraft Parts & Auxiliary Equipment, NEC
126	KHC	KRAFT HEINZ CO	31-Mar-1964	6-Jun-2013	19023	2030	Canned, Frozen & Preserved Fruit, Veg & Food Specialties
127	JCP	PENNEY (J C) CO	31-Mar-1964	1-Dec-2013	19023	5311	Retail-Department Stores
128	BEAM	BEAM INC	31-Mar-1964	30-Apr-2014	19023	2085	
129	MWV	MEADWESTVACO CORP	31-Mar-1964	1-Jul-2015	19023	2631	Paperboard Mills
130	AA.3	ALCOA INC	31-Mar-1964	31-Oct-2016	19023	3720	Aircraft & Parts

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
131	PBI	PITNEY BOWES INC	31-Mar-1964	28-Feb-2017	19023	3579	Office Machines, NEC
132	BHI	BAKER HUGHES INC	31-Mar-1964	4-Jul-2017	19023	1381	Drilling Oil & Gas Wells
133	DD	DU PONT (E I) DE NEMOURS	31-Mar-1964	31-Aug-2017	19023	2820	Plastic Material, Synth Resin/Rubber, Cellulos (No Glass)
134	SHW	SHERWIN-WILLIAMS CO	30-Jun-1964		18932	2851	Paints, Varnishes, Lacquers, Enamels & Allied Prods
135	TWX.1	TIME WARNER INC-OLD	30-Jun-1964	15-Jan-2001	18932	7812	Services-Motion Picture & Video Tape Production
136	CMI	CUMMINS INC	31-Mar-1965		18658	3510	Engines & Turbines
137	EMR	EMERSON ELECTRIC CO	31-Mar-1965		18658	3823	Industrial Instruments For Measurement, Display, and Control
138	SLB	SCHLUMBERGER LTD	31-Mar-1965		18658	1389	Oil & Gas Field Services, NEC
139	TMC	TIMES MIRROR CO -SER A	31-Dec-1965	12-Jun-2000	18383	2711	Newspapers: Publishing or Publishing & Printing
140	EFU.1	EASTERN ENTERPRISES	30-Sep-1966	1-Nov-2000	18110	4924	Natural Gas Distribution
141	TEK.1	TEKTRONIX INC	31-Mar-1967	15-Nov-2007	17928	3825	Instruments For Meas & Testing of Electricity & Elec Signals
142	AVP	AVON PRODUCTS	30-Jun-1967	22-Mar-2015	17837	2844	Perfumes, Cosmetics & Other Toilet Preparations
143	CSX	CSX CORP	30-Sep-1967		17745	4011	Railroads, Line-Haul Operating

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
144	SMI.1	SPRINGS INDUSTRIES -CL A	30-Sep-1967	11-Dec-2000	17745	2211	Broadwoven Fabric Mills, Cotton
145	DAL	DELTA AIR LINES INC	31-Mar-1968	18-Aug-2005	17562	4512	Air Transportation, Scheduled
146	BGG	BRIGGS & STRATTON	30-Jun-1968	2-Apr-2001	17471	3510	Engines & Turbines
147	MDP	MEREDITH CORP	30-Jun-1968	3-Jan-2011	17471	2721	Periodicals: Publishing or Publishing & Printing
148	CLX	CLOROX CO/DE	31-Mar-1969		17197	2842	Specialty Cleaning, Polishing and Sanitation Preparations
149	GIS	GENERAL MILLS INC	31-Mar-1969		17197	2040	Grain Mill Products
150	SUN.1	SUNOCO INC	31-Mar-1969	4-Oct-2012	17197	2911	Petroleum Refining
151	NEM	NEWMONT MINING CORP	30-Jun-1969		17106	1040	Gold and Silver Ores
152	4622B	CERIDIAN CORP	30-Jun-1969	1-Apr-2001	17106	8721	
153	MCD	MCDONALD'S CORP	30-Jun-1970		16741	5812	Retail-Eating Places
154	CHA.3	CHAMPION INTERNATIONAL CORP	30-Jun-1970	18-Jun-2000	16741	2621	Paper Mills
155	PCH	POTLATCHDELTIC CORP	30-Jun-1970	8-Jul-2001	16741	2421	Sawmills & Planing Mills, General
156	HLT	HILTON WORLDWIDE HOLDINGS	30-Jun-1970	24-Oct-2007	16741	7011	Hotels & Motels
157	OMX	OFFICEMAX INC	30-Jun-1970	22-Jun-2008	16741	5110	Wholesale-Paper & Paper Products
158	LLY	LILLY (ELI) & CO	31-Dec-1970		16557	2834	Pharmaceutical Preparations
159	ACV.1	ALBERTO-CULVER CO	31-Mar-1971	16-Nov-2006	16467	2844	Perfumes, Cosmetics & Other Toilet Preparations

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
160	MDR	MCDERMOTT INTL INC	30-Sep-1971	19-Aug-2003	16284	3730	Ship & Boat Building & Repairing
161	BAX	BAXTER INTERNATIONAL INC	30-Sep-1972		15918	2834	Pharmaceutical Preparations
162	BDX	BECTON DICKINSON & CO	30-Sep-1972		15918	3841	Surgical & Medical Instruments & Apparatus
163	BC	BRUNSWICK CORP	31-Mar-1973	22-Jun-2008	15736	3510	Engines & Turbines
164	RSHCQ	RS LEGACY CORP	31-Mar-1973	30-Jun-2011	15736	5731	Retail-Radio, TV & Consumer Electronics Stores
165	JNJ	JOHNSON & JOHNSON	30-Jun-1973		15645	2834	Pharmaceutical Preparations
166	GPC	GENUINE PARTS CO	31-Dec-1973		15461	5013	Wholesale-Motor Vehicle Supplies & New Parts
167	FLTWQ	FLEETWOOD ENTERPRISES INC	30-Jun-1974	30-Jan-2000	15280	3716	Motor Homes
168	HPQ	HP INC	31-Dec-1974		15096	3570	Computer & office Equipment
169	TNB	THOMAS & BETTS CORP	31-Dec-1974	2-Aug-2004	15096	3640	Electric Lighting & Wiring Equipment
170	LPX	LOUISIANA-PACIFIC CORP	31-Dec-1974	9-Nov-2006	15096	2400	Lumber & Wood Products (No Furniture)
171	ABI.3	APPLIED BIOSYSTEMS INC	31-Dec-1974	23-Nov-2008	15096	2835	In Vitro & In Vivo Diagnostic Substances
172	WMB	WILLIAMS COS INC	31-Mar-1975		15006	4922	Natural Gas Transmission
173	MKG	MALLINCKRODT INC	31-Mar-1975	17-Oct-2000	15006	3842	Orthopedic, Prosthetic & Surgical Appliances & Supplies

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
174	BCR	BARD (C.R.) INC	30-Jun-1975	2-Jan-2018	14915	3841	Surgical & Medical Instruments & Apparatus
175	RAD	RITE AID CORP	31-Dec-1975	26-Jul-2000	14731	5912	Retail-Drug Stores and Proprietary Stores
176	KRI	KNIGHT-RIDDER INC	31-Dec-1975	27-Jun-2006	14731	2711	Newspapers: Publishing or Publishing & Printing
177	TGNA	TEGNA INC	31-Dec-1975	1-Jun-2017	14731	4833	Television Broadcasting Stations
178	IFF	INTL FLAVORS & FRAGRANCES	31-Mar-1976		14640	2860	Industrial Organic Chemicals
179	BDK	BLACK & DECKER CORP	31-Mar-1976	14-Mar-2010	14640	3540	Metalworkg Machinery & Equipment
180	AET	AETNA INC	30-Jun-1976		14549	6324	Hospital & Medical Service Plans
181	AXP	AMERICAN EXPRESS CO	30-Jun-1976		14549	6141	Personal Credit Institutions
182	CI	CIGNA CORP	30-Jun-1976		14549	6324	Hospital & Medical Service Plans
183	JPM	JPMORGAN CHASE & CO	30-Jun-1976		14549	6020	
184	TAP	MOLSON COORS BREWING CO	30-Jun-1976		14549	2082	Malt Beverages
185	DIS	DISNEY (WALT) CO	30-Jun-1976		14549	4888	
186	DUK	DUKE ENERGY CORP	30-Jun-1976		14549	4931	Electric & Other Services Combined
187	NEE	NEXTERA ENERGY INC	30-Jun-1976		14549	4911	Electric Services
188	LNC	LINCOLN NATIONAL CORP	30-Jun-1976		14549	6311	Life Insurance
189	BAC	BANK OF AMERICA CORP	30-Jun-1976		14549	6020	
190	WFC	WELLS FARGO & CO	30-Jun-1976		14549	6020	

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
191	JPM.Z	MORGAN (J P) & CO	30-Jun-1976	1-Jan-2001	14549	6020	
192	AGC.1	AMERICAN GENERAL CORP	30-Jun-1976	29-Aug-2001	14549	6311	Life Insurance
193	SPC.2	ST PAUL COS	30-Jun-1976	1-Apr-2004	14549	6331	Fire, Marine & Casualty Insurance
194	JP.1	JEFFERSON-PILOT CORP	30-Jun-1976	2-Apr-2006	14549	6311	Life Insurance
195	MEL.3	MELLON FINANCIAL CORP	30-Jun-1976	1-Jul-2007	14549	6020	
196	SAF	SAFECO CORP	30-Jun-1976	22-Sep-2008	14549	6331	Fire, Marine & Casualty Insurance
197	BUD.2	ANHEUSER-BUSCH COS INC	30-Jun-1976	18-Nov-2008	14549	2082	Malt Beverages
198	CB.1	CHUBB CORP	30-Jun-1976	18-Jan-2016	14549	6331	Fire, Marine & Casualty Insurance
199	TGT	TARGET CORP	31-Dec-1976		14365	5331	Retail-Variety Stores
200	INTC	INTEL CORP	31-Dec-1976		14365	3674	Semiconductors & Related Devices
201	NSM.2	NATIONAL SEMICONDUCTOR CORP	31-Dec-1976	25-Sep-2011	14365	3674	Semiconductors & Related Devices
202	CTX.2	CENTEX CORP	31-Mar-1977	18-Aug-2009	14275	1531	Operative Builders
203	DJ	DOW JONES & CO INC	30-Jun-1978	13-Dec-2007	13819	2711	Newspapers: Publishing or Publishing & Printing
204	BMS	BEMIS CO INC	30-Sep-1978	4-Dec-2014	13727	2670	Converted Paper & Paperboard Prods (No Containers/Boxes)
205	TXT	TEXTRON INC	31-Dec-1978		13635	3721	Aircraft

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
206	ATI	ALLEGHENY TECHNOLOGIES INC	31-Dec-1978	1-Jul-2015	13635	3350	Rolling Drawing & Extruding of Nonferrous Metals
207	THC	TENET HEALTHCARE CORP	31-Mar-1979	17-Apr-2016	13545	8062	Services-General Medical & Surgical Hospitals, NEC
208	VFC	VF CORP	30-Jun-1979		13454	2320	Men's & Boys' Furnishings, Work Clothing, & Allied Garments
209	WBA	WALGREENS BOOTS ALLIANCE INC	31-Dec-1979		13270	5912	Retail-Drug Stores and Proprietary Stores
210	DNB	DUN & BRADSTREET CORP	31-Dec-1979	2-Oct-2000	13270	7323	
211	AIG	AMERICAN INTERNATIONAL GROUP	31-Mar-1980		13179	6331	Fire, Marine & Casualty Insurance
212	FLR	FLUOR CORP	31-Mar-1980		13179	1600	Heavy Construction Other Than Bldg Const - Contractors
213	PNU.1	PHARMACIA & UPJOHN INC	30-Jun-1980	2-Apr-2000	13088	2834	Pharmaceutical Preparations
214	FDX	FEDEX CORP	31-Dec-1980		12904	4513	Air Courier Services
215	PCAR	PACCAR INC	31-Dec-1980		12904	3711	Motor Vehicles & Passenger Car Bodies
216	NRTLQ	NORTEL NETWORKS CORP	31-Dec-1980	21-Jul-2002	12904	9995	Non-Operating Establishments
217	ADP	AUTOMATIC DATA PROCESSING	31-Mar-1981		12814	7374	Services-Computer Processing & Data Preparation
218	SFA.1	SCIENTIFIC-ATLANTA INC	31-Mar-1981	26-Feb-2006	12814	3663	Radio & TV Broadcasting & Communications Equipment

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
219	GWW	GRAINGER (W W) INC	30-Jun-1981		12723	5000	Wholesale-Durable Goods
220	MAS	MASCO CORP	30-Jun-1981		12723	3430	Heating Equip, Except Elec & Warm Air; & Plumbing Fixtures
221	DXC	DXC TECHNOLOGY COMPANY	30-Jun-1981	30-Nov-2015	12723	7370	Services-Computer Programming, Data Processing, Etc.
222	ADM	ARCHER-DANIELS-MIDLAND CO	29-Jul-1981		12694	2070	Fats & Oils
223	H.2	HARCOURT GENERAL INC	30-Sep-1981	27-Jun-2001	12631	2731	Books: Publishing or Publishing & Printing
224	HSH	HILLSHIRE BRANDS CO	30-Sep-1981	28-Jun-2012	12631	2013	Sausages & Other Prepared Meat Products
225	MAT	MATTEL INC	31-Mar-1982		12449	3942	Dolls & Stuffed Toys
226	WEN.2	WENDY'S INTERNATIONAL INC	31-Mar-1982	29-Sep-2008	12449	5812	Retail-Eating Places
227	BAC2	MERRILL LYNCH & CO INC	30-Jun-1982	1-Jan-2009	12358	6211	Security Brokers, Dealers & Flotation Companies
228	WMT	WALMART INC	31-Aug-1982		12296	5331	Retail-Variety Stores
229	SNA	SNAP-ON INC	30-Sep-1982		12266	3420	Cutlery, Handtools & General Hardware
230	SWK	STANLEY BLACK & DECKER INC	30-Sep-1982		12266	3540	Metalworkg Machinery & Equipment
231	ROH	ROHM AND HAAS CO	30-Sep-1982	1-Apr-2009	12266	2821	Plastic Materials, Synth Resins & Nonvulcan Elastomers
232	BF.B	BROWN FORMAN CORP	31-Oct-1982		12235	2085	

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
233	AAPL	APPLE INC	30-Nov-1982		12205	3663	Radio & TV Broadcasting & Communications Equipment
234	OXY	OCCIDENTAL PETROLEUM CORP	31-Dec-1982		12174	1311	Crude Petroleum & Natural Gas
235	R	RYDER SYSTEM INC	31-Dec-1982	18-Jun-2017	12174	7510	Services-Auto Rental & Leasing (No Drivers)
236	EC.2	ENGELHARD CORP	31-Jul-1983	5-Jun-2006	11962	2810	Industrial Inorganic Chemicals
237	CAG	CONAGRA BRANDS INC	31-Aug-1983		11931	2000	Food and Kindred Products
238	LB	L BRANDS INC	30-Sep-1983		11901	5621	Retail-Women's Clothing Stores
239	1231B	TOYS R US INC	30-Sep-1983	21-Jul-2005	11901	5945	Retail-Hobby, Toy & Game Shops
240	VZ	VERIZON COMMUNICATIONS INC	30-Nov-1983		11840	4812	Radiotelephone Communications
241	T	AT&T INC	30-Nov-1983		11840	4812	Radiotelephone Communications
242	BLS	BELLSOUTH CORP	30-Nov-1983	3-Jan-2007	11840	4813	Telephone Communications (No Radiotelephone)
243	Q.2	QWEST COMMUNICATION INTL INC	30-Nov-1983	31-Mar-2011	11840	4813	Telephone Communications (No Radiotelephone)
244	LOW	LOWE'S COMPANIES INC	29-Feb-1984		11749	5211	Retail-Lumber & Other Building Materials Dealers
245	PHM	PULTEGROUP INC	30-Apr-1984		11688	1531	Operative Builders
246	HES	HESS CORP	31-May-1984		11657	1311	Crude Petroleum & Natural Gas
247	AMD	ADVANCED MICRO DEVICES	30-Jun-1984	22-Sep-2013	11627	3674	Semiconductors & Related Devices

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
248	LMT	LOCKHEED MARTIN CORP	31-Jul-1984		11596	3760	Guided Missiles & Space Vehicles & Parts
249	RRD	DONNELLEY (R R) & SONS CO	31-Jul-1984	11-Dec-2012	11596	2750	Commercial Printing
250	HAS	HASBRO INC	30-Sep-1984		11535	3944	Games, Toys & Children's Vehicles (No Dolls & Bicycles)
251	BLL	BALL CORP	31-Oct-1984		11504	3411	Metal Cans
252	SMS.2	SHARED MEDICAL SYSTEMS CORP	31-Oct-1984	7-Jun-2000	11504	7373	Services-Computer Integrated Systems Design
253	KMG.1	KERR-MCGEE CORP	31-Oct-1984	10-Aug-2006	11504	1311	Crude Petroleum & Natural Gas
254	KATE	KATE SPADE & CO	31-Oct-1984	1-Dec-2008	11504	2300	Apparel & Other Finished Prods of Fabrics & Similar Matl
255	RML.1	RUSSELL CORP	30-Nov-1984	11-Dec-2000	11474	2253	Knit Outerwear Mills
256	ABS.1	ALBERTSON'S INC	30-Nov-1984	1-Jun-2006	11474	5411	Retail-Grocery Stores
257	ANDW	ANDREW CORP	30-Nov-1984	1-Oct-2006	11474	3357	Drawing & Insulating of Nonferrous Wire
258	NYT	NEW YORK TIMES CO -CL A	30-Nov-1984	19-Dec-2010	11474	2711	Newspapers: Publishing or Publishing & Printing
259	SVU	SUPERVALU INC	31-Jan-1985	30-Apr-2012	11412	5141	Wholesale-Groceries, General Line (merchandise)
260	RDC	ROWAN COMPANIES PLC	31-Jan-1985	18-Aug-2014	11412	1381	Drilling Oil & Gas Wells
261	APD	AIR PRODUCTS & CHEMICALS INC	30-Apr-1985		11323	2810	Industrial Inorganic Chemicals

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
262	NUE	NUCOR CORP	30-Apr-1985		11323	3312	Steel Works, Blast Furnaces & Rolling Mills (Coke Ovens)
263	PKI	PERKINELMER INC	31-May-1985		11292	3826	Laboratory Analytical Instruments
264	EP2	EL PASO CGP CO	31-May-1985	29-Jan-2001	11292	4922	Natural Gas Transmission
265	NOC	NORTHROP GRUMMAN CORP	30-Jun-1985		11262	3812	Search, Detection, Navigation, Guidance, Aeronautical Sys
266	CNP	CENTERPOINT ENERGY INC	31-Jul-1985		11231	4931	Electric & Other Services Combined
267	TJX	TJX COMPANIES INC	30-Sep-1985		11170	5651	Retail-Family Clothing Stores
268	DOV	DOVER CORP	31-Oct-1985		11139	3585	Air-Cond & Warm Air Heatg Equip & Comm & Indl Refrig Equip
269	PH	PARKER-HANNIFIN CORP	30-Nov-1985		11109	3490	Miscellaneous Fabricated Metal Products
270	LDG	LONGS DRUG STORES CORP	30-Nov-1985	1-Jul-2001	11109	5912	Retail-Drug Stores and Proprietary Stores
271	TRCO	TRIBUNE MEDIA CO	31-Jan-1986	20-Dec-2007	11047	4833	Television Broadcasting Stations
272	ITW	ILLINOIS TOOL WORKS	28-Feb-1986		11019	3560	General Industrial Machinery & Equipment
273	DLX	DELUXE CORP	30-Apr-1986	19-Dec-2004	10958	2780	Blankbooks, Looseleaf Binders & Bookbinding & Related Work

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
274	JCI	JOHNSON CONTROLS INTL PLC	31-May-1986		10927	3585	Air-Cond & Warm Air Heatg Equip & Comm & Indl Refrig Equip
275	FJ	FORT JAMES CORP	31-May-1986	23-Nov-2000	10927	2621	Paper Mills
276	6583B	BAUSCH & LOMB HLDGS -REDH	31-May-1986	28-Oct-2007	10927	2834	Pharmaceutical Preparations
277	GPS	GAP INC	31-Aug-1986		10835	5651	Retail-Family Clothing Stores
278	JWN	NORDSTROM INC	31-Aug-1986		10835	5651	Retail-Family Clothing Stores
279	MDT	MEDTRONIC PLC	31-Oct-1986		10774	3845	Electromedical & Electrotherapeutic Apparatus
280	DDS	DILLARDS INC -CL A	31-Oct-1986	21-Oct-2008	10774	5311	Retail-Department Stores
281	HRB	BLOCK H & R INC	30-Nov-1986		10744	7200	Services-Personal Services
282	SYY	SYSCO CORP	31-Dec-1986		10713	5140	Wholesale-Groceries & Related Products
283	SCI	SERVICE CORP INTERNATIONAL	31-Dec-1986	15-Mar-2000	10713	7200	Services-Personal Services
284	NSI.1	NATIONAL SERVICE INDS INC	31-Dec-1986	2-Dec-2001	10713	7200	Services-Personal Services
285	3UAIRQ	US AIRWAYS GROUP INC-OLD	31-Dec-1986	14-May-2002	10713	4512	Air Transportation, Scheduled
286	TIN	TEMPLE-INLAND INC	31-Dec-1986	30-Dec-2007	10713	2631	Paperboard Mills
287	RBK	REEBOK INTERNATIONAL LTD	31-Jan-1987	31-Jan-2006	10682	3021	Rubber & Plastics Footwear
288	UST.1	UST INC	30-Apr-1987	5-Jan-2009	10593	2100	Tobacco Products
289	CA	CA INC	31-Jul-1987		10501	7372	Services-Prepackaged Software
290	MMC	MARSH & MCLENNAN COS	31-Aug-1987		10470	6411	Insurance Agents, Brokers & Service

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
291	PDG	PLACER DOME INC	31-Aug-1987	21-Jul-2002	10470	1040	Gold and Silver Ores
292	AVY	AVERY DENNISON CORP	31-Dec-1987		10348	2670	Converted Paper & Paperboard Prods (No Containers/Boxes)
293	PLL	PALL CORP	31-Dec-1987	30-Aug-2015	10348	3569	General Industrial Machinery & Equipment, NEC
294	IKN	IKON OFFICE SOLUTIONS	29-Feb-1988	28-Jun-2000	10288	5040	Wholesale-Professional & Commercial Equipment & Supplies
295	CPQ.2	COMPAQ COMPUTER CORP	29-Feb-1988	5-May-2002	10288	3571	Electronic Computers
296	HD	HOME DEPOT INC	31-Mar-1988		10257	5211	Retail-Lumber & Other Building Materials Dealers
297	PNC	PNC FINANCIAL SVCS GROUP INC	30-Apr-1988		10227	6020	
298	ONE.1	BANK ONE CORP	30-Apr-1988	30-Jun-2004	10227	6020	
299	C	CITIGROUP INC	31-May-1988		10196	6199	Finance Services
300	STI	SUNTRUST BANKS INC	31-May-1988		10196	6020	
301	GDW	GOLDEN WEST FINANCIAL CORP	30-Jun-1988	1-Oct-2006	10166	6035	Savings Institution, Federally Chartered
302	WOR	WORTHINGTON INDUSTRIES	31-Jul-1988	19-Dec-2004	10135	3310	Steel Works, Blast Furnaces & Rolling & Finishing Mills
303	MIL.2	MILLIPORE CORP	31-Jul-1988	14-Jul-2010	10135	3826	Laboratory Analytical Instruments

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
304	FNMA	FANNIE MAE	31-Aug-1988	10-Sep-2008	10104	6111	Federal & Federally Sponsored Credit Agencies
305	NKE	NIKE INC	30-Nov-1988		10013	3021	Rubber & Plastics Footwear
306	FBF	FLEETBOSTON FINANCIAL CORP	30-Nov-1988	31-Mar-2004	10013	6020	
307	ECL	ECOLAB INC	31-Jan-1989		9951	2842	Specialty Cleaning, Polishing and Sanitation Preparations
308	WB.3	WACHOVIA CORP	28-Feb-1989	1-Jan-2009	9923	6020	
309	KBH	KB HOME	31-Mar-1989	20-Dec-2009	9892	1531	Operative Builders
310	NWL	NEWELL BRANDS INC	30-Apr-1989		9862	3990	Miscellaneous Manufacturing Industries
311	TMK	TORCHMARK CORP	30-Apr-1989		9862	6311	Life Insurance
312	CCTYQ	CIRCUIT CITY STORES INC	31-May-1989	30-Mar-2008	9831	5731	Retail-Radio, TV & Consumer Electronics Stores
313	JOSEA	JOSTENS INC	31-Jul-1989	9-May-2000	9770	3911	Jewelry, Precious Metal
314	TYC	TYCO INTERNATIONAL PLC	31-Jul-1989	16-Mar-2009	9770	3669	Communications Equipment, NEC
315	S	SPRINT CORP	31-Jul-1989	8-Jul-2013	9770	4812	Radiotelephone Communications
316	ORCL	ORACLE CORP	31-Aug-1989		9739	7372	Services-Prepackaged Software
317	ASH	ASHLAND GLOBAL HOLDINGS INC	31-Aug-1989	13-Nov-2008	9739	2820	Plastic Material, Synth Resin/Rubber, Cellulos (No Glass)
318	GAS.2	NICOR INC	31-Aug-1989	12-Dec-2011	9739	4924	Natural Gas Distribution

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
319	DUK8	PROGRESS ENERGY INC	31-Aug-1989	1-Jul-2012	9739	4911	Electric Services
320	AM.1	AMERICAN GREETINGS -CL A	31-Oct-1989	2-May-2004	9678	2771	Greeting Cards
321	STJ	ST JUDE MEDICAL INC	30-Nov-1989	4-Jan-2017	9648	3845	Electromedical & Electrotherapeutic Apparatus
322	ADSK	AUTODESK INC	1-Dec-1989		9647	7370	Services-Computer Programming, Data Processing, Etc.
323	CZR	CAESARS ENTERTAINMENT CORP	1-Feb-1990	28-Jan-2008	9585	7990	Services-Miscellaneous Amusement & Recreation
324	CTB	COOPER TIRE & RUBBER CO	1-May-1990	17-Jul-2006	9496	3011	Tires & Inner Tubes
325	PBY	PEP BOYS-MANNY MOE & JACK	1-Aug-1990	2-Apr-2000	9404	5531	Retail-Auto & Home Supply Stores
326	5938B	BIOMET INC	1-Aug-1990	11-Jul-2007	9404	3842	Orthopedic, Prosthetic & Surgical Appliances & Supplies
327	AZA.1	ALZA CORP	11-Dec-1990	24-Jun-2001	9272	2834	Pharmaceutical Preparations
328	MRO	MARATHON OIL CORP	1-May-1991		9131	1311	Crude Petroleum & Natural Gas
329	X	UNITED STATES STEEL CORP	1-May-1991	1-Jul-2014	9131	3312	Steel Works, Blast Furnaces & Rolling Mills (Coke Ovens)
330	AEE	AMEREN CORP	19-Sep-1991		8990	4931	Electric & Other Services Combined
331	NOVL	NOVELL INC	1-Oct-1991	27-Apr-2011	8978	7370	Services-Computer Programming, Data Processing, Etc.
332	GLK.1	GREAT LAKES CHEMICAL CORP	1-Nov-1991	4-Jul-2005	8947	2890	Miscellaneous Chemical Products

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
333	AMGN	AMGEN INC	2-Jan-1992		8885	2836	Biological Products, (No Diagnostic Substances)
334	FMCC	FEDERAL HOME LOAN MORTG CORP	2-Jan-1992	10-Sep-2008	8885	6111	Federal & Federally Sponsored Credit Agencies
335	AGN.2	ALLERGAN INC	7-Feb-1992	22-Mar-2015	8849	2834	Pharmaceutical Preparations
336	KRB	MBNA CORP	1-Apr-1992	2-Jan-2006	8795	6020	
337	PX	PRAXAIR INC	1-Jul-1992		8704	2810	Industrial Inorganic Chemicals
338	JAVA	SUN MICROSYSTEMS INC	1-Aug-1992	28-Jan-2010	8673	3571	Electronic Computers
339	IPG	INTERPUBLIC GROUP OF COS	1-Oct-1992		8612	7311	Services-Advertising Agencies
340	RAL.	RALSTON PURINA CO	1-Aug-1993	12-Dec-2001	8308	2040	Grain Mill Products
341	HCA	HCA HEALTHCARE INC	1-Sep-1993	19-Nov-2006	8277	8062	Services-General Medical & Surgical Hospitals, NEC
342	COST	COSTCO WHOLESALE CORP	1-Oct-1993		8247	5399	Retail-Misc General Merchandise Stores
343	WB.1	WACHOVIA CORP-OLD	1-Oct-1993	3-Sep-2001	8247	6020	
344	ABX	BARRICK GOLD CORP	1-Nov-1993	21-Jul-2002	8216	1040	Gold and Silver Ores
345	BR.2	BURLINGTON RESOURCES INC	1-Nov-1993	2-Apr-2006	8216	1311	Crude Petroleum & Natural Gas
346	CSCO	CISCO SYSTEMS INC	1-Dec-1993		8186	3576	Computer Communications Equipment
347	CMCSA	COMCAST CORP	1-Dec-1993	18-Nov-2002	8186	4841	Cable & Other Pay Television Services

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
348	EMN	EASTMAN CHEMICAL CO	1-Jan-1994		8155	2821	Plastic Materials, Synth Resins & Nonvulcan Elastomers
349	KEY	KEYCORP	1-Mar-1994		8096	6020	
350	UNM	UNUM GROUP	1-Mar-1994		8096	6321	Accident & Health Insurance
351	MSFT	MICROSOFT CORP	1-Jun-1994		8004	7372	Services-Prepackaged Software
352	LUV	SOUTHWEST AIRLINES	1-Jul-1994		7974	4512	Air Transportation, Scheduled
353	UNH	UNITEDHEALTH GROUP INC	1-Jul-1994		7974	6324	Hospital & Medical Service Plans
354	CBS	CBS CORP	1-Sep-1994		7912	4888	
355	SIAL	SIGMA-ALDRICH CORP	1-Sep-1994	18-Nov-2015	7912	2833	Medicinal Chemicals & Botanical Products
356	MU	MICRON TECHNOLOGY INC	27-Sep-1994		7886	3674	Semiconductors & Related Devices
357	FDC	FIRST DATA CORP	27-Sep-1994	24-Sep-2007	7886	7374	Services-Computer Processing & Data Preparation
358	NCC	NATIONAL CITY CORP	27-Sep-1994	1-Jan-2009	7886	6020	
359	DUK6	CINERGY CORP	25-Oct-1994	2-Apr-2006	7858	4911	Electric Services
360	AT.2	ALLTEL CORP	21-Dec-1994	18-Nov-2007	7801	4812	Radiotelephone Communications
361	SGICQ	SILICON GRAPHICS INC	18-Jan-1995	20-Jun-2000	7773	7370	Services-Computer Programming, Data Processing, Etc.
362	GPU.	GPU INC	9-Feb-1995	6-Nov-2001	7751	4911	Electric Services
363	BSX	BOSTON SCIENTIFIC CORP	24-Feb-1995		7736	3845	Electromedical & Electrotherapeutic Apparatus

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
364	AMAT	APPLIED MATERIALS INC	16-Mar-1995		7716	3559	Special Industry Machinery, NEC
365	BK	BANK OF NEW YORK MELLON CORP	31-Mar-1995		7701	6020	
366	CAR	AVIS BUDGET GROUP INC	1-May-1995	31-Jul-2006	7670	7510	Services-Auto Rental & Leasing (No Drivers)
367	L	LOEWS CORP	10-May-1995		7661	6331	Fire, Marine & Casualty Insurance
368	ETS	ENTERASYS NETWORKS INC	22-May-1995	5-Aug-2001	7649	3576	Computer Communications Equipment
369	DRI	DARDEN RESTAURANTS INC	30-May-1995		7641	5812	Retail-Eating Places
370	TLAB	TELLABS INC	3-Jul-1995	20-Dec-2011	7607	3661	Telephone & Telegraph Apparatus
371	HBC2	HSBC USA INC	5-Jul-1995	2-Jan-2000	7605	6020	
372	ALL	ALLSTATE CORP	13-Jul-1995		7597	6331	Fire, Marine & Casualty Insurance
373	FCX	FREEPORT-MCMORAN INC	31-Jul-1995		7579	1000	Metal Mining
374	WLL.1	WILLAMETTE INDUSTRIES	18-Aug-1995	10-Feb-2002	7561	2621	Paper Mills
375	MS	MORGAN STANLEY	22-Sep-1995		7526	6211	Security Brokers, Dealers & Flotation Companies
376	USB.1	U S BANCORP/DE-OLD	23-Oct-1995	26-Feb-2001	7495	6020	
377	UMG.Z	MEDIAONE GROUP INC	1-Nov-1995	15-Jun-2000	7486	4841	Cable & Other Pay Television Services

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
378	M	MACY'S INC	8-Nov-1995		7479	5311	Retail-Department Stores
379	PPL	PPL CORP	27-Nov-1995		7460	4911	Electric Services
380	CMA	COMERICA INC	1-Dec-1995		7456	6020	
381	HUM	HUMANA INC	1-Dec-1995		7456	6324	Hospital & Medical Service Plans
382	LSI.1	LSI CORP	14-Dec-1995	7-May-2014	7443	3674	Semiconductors & Related Devices
383	HIG	HARTFORD FINANCIAL SERVICES	20-Dec-1995		7437	6331	Fire, Marine & Casualty Insurance
384	COMS	3COM CORP	2-Jan-1996	27-Jul-2000	7424	3576	Computer Communications Equipment
385	FITB	FIFTH THIRD BANCORP	8-Mar-1996		7358	6020	
386	EMC	EMC CORP/MA	28-Mar-1996	7-Sep-2016	7338	3572	Computer Storage Devices
387	GIC.3	GENERAL INSTRUMENT CORP	1-Apr-1996	5-Jan-2000	7334	3663	Radio & TV Broadcasting & Communications Equipment
388	MCIP	MCI INC	1-Apr-1996	14-May-2002	7334	4813	Telephone Communications (No Radiotelephone)
389	AON	AON PLC	23-Apr-1996		7312	6411	Insurance Agents, Brokers & Service
390	TUP	TUPPERWARE BRANDS CORP	31-May-1996	24-Mar-2004	7274	3089	Plastics Products, NEC
391	MTG	MGIC INVESTMENT CORP/WI	19-Jul-1996	30-Oct-2008	7225	6351	Surety Insurance
392	SEG.2	SEAGATE TECHNOLOGY-OLD	16-Aug-1996	21-Nov-2000	7197	3572	Computer Storage Devices
393	7732B	DELL TECHNOLOGIES INC	6-Sep-1996	28-Oct-2013	7176	3571	Electronic Computers

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
394	LU	LUCENT TECHNOLOGIES INC	1-Oct-1996	30-Nov-2006	7151	7373	Services-Computer Integrated Systems Design
395	UPR.1	UNION PACIFIC RESOURCES GRP	16-Oct-1996	16-Jul-2000	7136	1311	Crude Petroleum & Natural Gas
396	IMS	IMS HEALTH HOLDINGS INC	4-Nov-1996	28-Feb-2010	7117	8700	Services-Engineering, Accounting, Research, Management
397	MBI	MBIA INC	3-Dec-1996	20-Dec-2009	7088	6351	Surety Insurance
398	GDT	GUIDANT CORP	19-Dec-1996	23-Apr-2006	7072	3841	Surgical & Medical Instruments & Apparatus
399	TMO	THERMO FISHER SCIENTIFIC INC	2-Jan-1997		7058	3826	Laboratory Analytical Instruments
400	AZO	AUTOZONE INC	2-Jan-1997		7058	5531	Retail-Auto & Home Supply Stores
401	EHC	ENCOMPASS HEALTH CORP	8-Jan-1997	20-Mar-2003	7052	8060	Services-Hospitals
402	CNO	CNO FINANCIAL GROUP INC	15-Jan-1997	24-Jul-2002	7045	6321	Accident & Health Insurance
403	PTC	PTC INC	3-Apr-1997	2-Jan-2007	6967	7372	Services-Prepackaged Software
404	ADBE	ADOBE SYSTEMS INC	6-May-1997		6934	7372	Services-Prepackaged Software
405	CAH	CARDINAL HEALTH INC	27-May-1997		6913	5122	Wholesale-Drugs, Proprietaries & Druggists' Sundries
406	SCHW	SCHWAB (CHARLES) CORP	2-Jun-1997		6907	6282	Investment Advice
407	PVN	PROVIDIAN FINANCIAL CORP	11-Jun-1997	2-Oct-2005	6898	6020	
408	CFC.3	COUNTRYWIDE FINANCIAL CORP	18-Jun-1997	30-Jun-2008	6891	6162	Mortgage Bankers & Loan Correspondents

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
409	EFX	EQUIFAX INC	19-Jun-1997		6890	7323	
410	WAMUQ	WASHINGTON MUTUAL INC	2-Jul-1997	29-Sep-2008	6877	6035	Savings Institution, Federally Chartered
411	APA	APACHE CORP	28-Jul-1997		6851	1311	Crude Petroleum & Natural Gas
412	APC	ANADARKO PETROLEUM CORP	28-Jul-1997		6851	1311	Crude Petroleum & Natural Gas
413	PGR	PROGRESSIVE CORP-OHIO	4-Aug-1997		6844	6331	Fire, Marine & Casualty Insurance
414	OI	OWENS-ILLINOIS INC	4-Aug-1997	10-Dec-2000	6844	3221	Glass Containers
415	HBAN	HUNTINGTON BANCSHARES	6-Aug-1997		6842	6020	
416	MIR.1	MIRAGE RESORTS INC	7-Aug-1997	31-May-2000	6841	7990	Services-Miscellaneous Amusement & Recreation
417	STT	STATE STREET CORP	18-Aug-1997		6830	6020	
418	IHRTQ	IHEARTMEDIA INC	1-Sep-1997	30-Jul-2008	6816	4832	Radio Broadcasting Stations
419	KLAC	KLA-TENCOR CORP	15-Sep-1997		6802	3827	Optical Instruments & Lenses
420	YUM	YUM BRANDS INC	7-Oct-1997		6780	5812	Retail-Eating Places
421	SNV	SYNOVUS FINANCIAL CORP	28-Nov-1997	1-Jan-2008	6728	6020	
422	BBT	BB&T CORP	4-Dec-1997		6722	6020	
423	CINF	CINCINNATI FINANCIAL CORP	18-Dec-1997		6708	6331	Fire, Marine & Casualty Insurance
424	OMC	OMNICOM GROUP	24-Dec-1997		6702	7311	Services-Advertising Agencies
425	LEHMQ	LEHMAN BROTHERS HOLDINGS INC	12-Jan-1998	16-Sep-2008	6683	6211	Security Brokers, Dealers & Flotation Companies

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
426	BIG	BIG LOTS INC	20-Jan-1998	13-Feb-2013	6675	5331	Retail-Variety Stores
427	SUB.2	SUMMIT BANCORP	28-Jan-1998	28-Feb-2001	6667	6020	
428	NTRS	NORTHERN TRUST CORP	30-Jan-1998		6665	6020	
429	SEE	SEALED AIR CORP	1-Apr-1998		6604	3081	Unsupported Plastics Film & Sheet
430	NXTL	NEXTEL COMMUNICATIONS INC	1-Apr-1998	14-Aug-2005	6604	4812	Radiotelephone Communications
431	AFS	ASSOCIATES FIRST CAP -CL A	8-Apr-1998	3-Dec-2000	6597	6141	Personal Credit Institutions
432	GTW	GATEWAY INC	27-Apr-1998	31-Jul-2006	6578	3571	Electronic Computers
433	BEN	FRANKLIN RESOURCES INC	29-Apr-1998		6576	6282	Investment Advice
434	MAR	MARRIOTT INTL INC	21-May-1998		6554	7011	Hotels & Motels
435	COF	CAPITAL ONE FINANCIAL CORP	1-Jul-1998		6513	6141	Personal Credit Institutions
436	BSC.1	BEAR STEARNS COMPANIES INC	1-Jul-1998	1-Jun-2008	6513	6211	Security Brokers, Dealers & Flotation Companies
437	NAVI	NAVIENT CORP	1-Jul-1998	4-Jun-2018	6513	6111	Federal & Federally Sponsored Credit Agencies
438	DG	DOLLAR GENERAL CORP	16-Jul-1998	8-Jul-2007	6498	5331	Retail-Variety Stores
439	WM	WASTE MANAGEMENT INC	17-Jul-1998		6497	4953	Refuse Systems
440	EDS.	ELECTRONIC DATA SYSTEMS CORP	11-Aug-1998	26-Aug-2008	6472	7370	Services-Computer Programming, Data Processing, Etc.
441	KSS	KOHL'S CORP	14-Aug-1998		6469	5311	Retail-Department Stores
442	RF	REGIONS FINANCIAL CORP	28-Aug-1998		6455	6020	

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
443	NGH.2	NABISCO GROUP HOLDINGS CORP	16-Sep-1998	11-Dec-2000	6436	2052	Cookies & Crackers
444	HCR	MANOR CARE INC	28-Sep-1998	8-Nov-2007	6424	8051	Services-Skilled Nursing Care Facilities
445	PAYX	PAYCHEX INC	1-Oct-1998		6421	8721	
446	UPC.	UNION PLANTERS CORP	1-Oct-1998	30-Jun-2004	6421	6020	
447	BMC	BMC SOFTWARE INC	1-Oct-1998	10-Sep-2013	6421	7372	Services-Prepackaged Software
448	AES	AES CORP	2-Oct-1998		6420	4991	Co-generation Services & Small Power Producers
449	PSFT.	PEOPLESOFT INC	2-Oct-1998	28-Dec-2004	6420	7372	Services-Prepackaged Software
450	SPLS	STAPLES INC	7-Oct-1998	17-Sep-2017	6415	5110	Wholesale-Paper & Paper Products
451	CCE	COCA-COLA EUROPEAN PARTNERS	8-Oct-1998	30-May-2016	6414	2086	Bottled & Canned Soft Drinks & Carbonated Waters
452	NCE.1	NEW CENTURY ENERGIES INC	3-Nov-1998	20-Aug-2000	6388	4931	Electric & Other Services Combined
453	SWY	SAFEWAY INC	13-Nov-1998	26-Jan-2015	6378	5411	Retail-Grocery Stores
454	DHR	DANAHER CORP	18-Nov-1998		6373	3826	Laboratory Analytical Instruments
455	PCS.1	SPRINT PCS GROUP	24-Nov-1998	22-Apr-2004	6367	4812	Radiotelephone Communications
456	CCL	CARNIVAL CORP/PLC (USA)	22-Dec-1998		6339	4400	Water Transportation
457	SLR	SOLECTRON CORP	31-Dec-1998	1-Oct-2007	6330	3672	Printed Circuit Boards

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
458	USB	U S BANCORP	4-Jan-1999		6326	6020	
459	CPWR.1	COMPUWARE CORP	4-Jan-1999	2-Jan-2012	6326	7372	Services-Prepackaged Software
460	TWX	TIME WARNER INC	4-Jan-1999	19-Jun-2018	6326	4888	
461	MCK	MCKESSON CORP	13-Jan-1999		6317	5122	Wholesale-Drugs, Proprietaries & Druggists' Sundries
462	SOTR	SOUTHTRUST CORP	1-Mar-1999	31-Oct-2004	6270	6020	
463	ASO.1	AMSOUTH BANCORPORATION	10-Mar-1999	5-Nov-2006	6261	6020	
464	CTL	CENTURYLINK INC	25-Mar-1999		6246	4813	Telephone Communications (No Radiotelephone)
465	KSU	KANSAS CITY SOUTHERN	5-Apr-1999	12-Jul-2000	6235	4011	Railroads, Line-Haul Operating
466	AGN	ALLERGAN PLC	12-Apr-1999		6228	2834	Pharmaceutical Preparations
467	CMS	CMS ENERGY CORP	3-May-1999		6207	4931	Electric & Other Services Combined
468	AFL	AFLAC INC	28-May-1999		6182	6321	Accident & Health Insurance
469	APTV	APTIV PLC	28-May-1999	10-Oct-2005	6182	3714	Motor Vehicle Parts & Accessories
470	PWJ.	PAINE WEBBER GROUP	4-Jun-1999	5-Nov-2000	6175	6211	Security Brokers, Dealers & Flotation Companies
471	WLP.1	WELLPOINT HEALTH NETWRKS INC	9-Jun-1999	30-Nov-2004	6170	6324	Hospital & Medical Service Plans
472	PGN3	FLORIDA PROGRESS CORP	22-Jun-1999	30-Nov-2000	6157	4911	Electric Services

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
473	ODP	OFFICE DEPOT INC	24-Jun-1999	19-Dec-2010	6155	5940	Retail-Miscellaneous Shopping Goods Stores
474	NTAP	NETAPP INC	25-Jun-1999		6154	3572	Computer Storage Devices
475	BBY	BEST BUY CO INC	30-Jun-1999		6149	5731	Retail-Radio, TV & Consumer Electronics Stores
476	VMC	VULCAN MATERIALS CO	1-Jul-1999		6148	1400	Mining & Quarrying of Nonmetallic Minerals (No Fuels)
477	QCOM	QUALCOMM INC	22-Jul-1999		6127	3674	Semiconductors & Related Devices
478	ADCT	ADC TELECOMMUNICATIONS INC	2-Aug-1999	1-Jul-2007	6116	3661	Telephone & Telegraph Apparatus
479	AW	ALLIED WASTE INDUSTRIES INC	2-Aug-1999	4-Dec-2008	6116	4953	Refuse Systems
480	COC1	CONOCO INC	9-Aug-1999	2-Sep-2002	6109	2911	Petroleum Refining
481	LXK	LEXMARK INTL INC -CL A	13-Aug-1999	30-Sep-2012	6105	3577	Computer Peripheral Equipment, NEC
482	TOS.1	TOSCO CORP	20-Sep-1999	17-Sep-2001	6067	2911	Petroleum Refining
483	GLBC	GLOBAL CROSSING LTD	29-Sep-1999	9-Oct-2001	6058	4813	Telephone Communications (No Radiotelephone)
484	SXCL	STEEL EXCEL INC	1-Oct-1999	13-May-2001	6056	1389	Oil & Gas Field Services, NEC
485	BBBY	BED BATH & BEYOND INC	1-Oct-1999	25-Jul-2017	6056	5700	Retail-Home Furniture, Furnishings & Equipment Stores
486	PNW	PINNACLE WEST CAPITAL CORP	4-Oct-1999		6053	4911	Electric Services

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
487	ADI	ANALOG DEVICES	12-Oct-1999		6045	3674	Semiconductors & Related Devices
488	TROW	PRICE (T. ROWE) GROUP	13-Oct-1999		6044	6282	Investment Advice
489	LEG	LEGGETT & PLATT INC	18-Oct-1999		6039	2510	Household Furniture
490	EP	EL PASO CORP	26-Oct-1999	24-May-2012	6031	4922	Natural Gas Transmission
491	CMVT	COMVERSE TECHNOLOGY INC	27-Oct-1999	31-Jan-2007	6030	7372	Services-Prepackaged Software
492	PTV	PACTIV CORP	5-Nov-1999	16-Nov-2010	6021	3089	Plastics Products, NEC
493	XLNX	XILINX INC	8-Nov-1999		6018	3674	Semiconductors & Related Devices
494	TER	TERADYNE INC	15-Nov-1999	22-Dec-2013	6011	3825	Instruments For Meas & Testing of Electricity & Elec Signals
495	0573B	QUINTILES TRANSNATIONAL CORP	16-Nov-1999	25-Sep-2003	6010	8731	Services-Commercial Physical & Biological Research
496	CTXS	CITRIX SYSTEMS INC	1-Dec-1999		5995	7372	Services-Prepackaged Software
497	MOLX	MOLEX INC	1-Dec-1999	9-Dec-2013	5995	3678	Electronic Connectors
498	OK.2	OLD KENT FINANCIAL CORP	2-Dec-1999	1-Apr-2001	5994	6020	
499	AABA	ALTABA INC	8-Dec-1999	18-Jun-2017	5988	7370	Services-Computer Programming, Data Processing, Etc.
500	RIG	TRANSOCEAN LTD	31-Dec-1999	18-Dec-2008	5965	1381	Drilling Oil & Gas Wells
501	NCR	NCR CORP	4-Jan-2000	30-Sep-2007	5961	7370	Services-Computer Programming, Data Processing, Etc.
502	YNR.1	YOUNG & RUBICAM INC	6-Jan-2000	2-Oct-2000	5959	7311	Services-Advertising Agencies

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
503	HOG	HARLEY-DAVIDSON INC	31-Jan-2000		5934	3751	Motorcycles, Bicycles & Parts
504	CNXT.1	CONEXANT SYSTEMS INC	31-Jan-2000	25-Jun-2002	5934	3674	Semiconductors & Related Devices
505	BGEN	BIOGEN INC-OLD	31-Jan-2000	12-Nov-2003	5934	2836	Biological Products, (No Diagnostic Substances)
506	TSG.2	SABRE HOLDINGS CORP -CL A	16-Mar-2000	1-Apr-2007	5889	7373	Services-Computer Integrated Systems Design
507	VRTS.1	VERITAS SOFTWARE CORP	3-Apr-2000	4-Jul-2005	5871	7372	Services-Prepackaged Software
508	LLTC	LINEAR TECHNOLOGY CORP	3-Apr-2000	12-Mar-2017	5871	3674	Semiconductors & Related Devices
509	ALTR.1	ALTERA CORP	18-Apr-2000	28-Dec-2015	5856	3674	Semiconductors & Related Devices
510	SAPE	SAPIENT CORP	5-May-2000	12-May-2002	5839	7373	Services-Computer Integrated Systems Design
511	SEBL	SIEBEL SYSTEMS INC	5-May-2000	31-Jan-2006	5839	7372	Services-Prepackaged Software
512	MXIM	MAXIM INTEGRATED PRODUCTS	10-May-2000	26-Sep-2007	5834	3674	Semiconductors & Related Devices
513	APCC.	AMERICAN POWER CONVERSION CP	1-Jun-2000	14-Feb-2007	5812	3620	Electrical Industrial Apparatus
514	A	AGILENT TECHNOLOGIES INC	5-Jun-2000		5808	3826	Laboratory Analytical Instruments
515	SBUX	STARBUCKS CORP	8-Jun-2000		5805	5812	Retail-Eating Places

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
516	CVG	CONVERGYS CORP	13-Jun-2000	20-Dec-2009	5800	7389	Services-Business Services, NEC
517	MEDI	MEDIMMUNE INC	16-Jun-2000	31-May-2007	5797	2836	Biological Products, (No Diagnostic Substances)
518	CF.6	CHARTER ONE FINANCIAL INC	19-Jun-2000	31-Aug-2004	5794	6020	
519	NVLS.1	NOVELLUS SYSTEMS INC	19-Jun-2000	4-Jun-2012	5794	3559	Special Industry Machinery, NEC
520	TIF	TIFFANY & CO	21-Jun-2000		5792	5944	Retail-Jewelry Stores
521	SANM	SANMINA CORP	21-Jun-2000	1-Jul-2007	5792	3672	Printed Circuit Boards
522	VC	VISTEON CORP	29-Jun-2000	2-Jan-2006	5784	3714	Motor Vehicle Parts & Accessories
523	MERQ	MERCURY INTERACTIVE CORP	29-Jun-2000	3-Jan-2006	5784	7372	Services-Prepackaged Software
524	BRCM	BROADCOM CORP	3-Jul-2000	31-Jan-2016	5780	3674	Semiconductors & Related Devices
525	JNS	JANUS CAPITAL GROUP INC	13-Jul-2000	22-Nov-2011	5770	6282	Investment Advice
526	CIT.1	CIT GROUP INC-OLD	17-Jul-2000	3-Jun-2001	5766	6172	Finance Lessors
527	VIAV	VIAVI SOLUTIONS INC	27-Jul-2000	22-Dec-2013	5756	3576	Computer Communications Equipment
528	PALM	PALM INC	28-Jul-2000	13-Aug-2002	5755	3571	Electronic Computers
529	KSE	KEYSPAN CORP	21-Aug-2000	26-Aug-2007	5731	4931	Electric & Other Services Combined
530	DVN	DEVON ENERGY CORP	30-Aug-2000		5722	1311	Crude Petroleum & Natural Gas
531	5933B	AVAYA INC	2-Oct-2000	25-Oct-2007	5689	3663	Radio & TV Broadcasting & Communications Equipment

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
532	MCO	MOODY'S CORP	3-Oct-2000		5688	7323	
533	DYN	DYNEGY INC	3-Oct-2000	20-Dec-2009	5688	1311	Crude Petroleum & Natural Gas
534	KG	KING PHARMACEUTICALS INC	3-Oct-2000	19-Dec-2010	5688	2834	Pharmaceutical Preparations
535	NBR	NABORS INDUSTRIES LTD	18-Oct-2000	22-Mar-2015	5673	1381	Drilling Oil & Gas Wells
536	PWER	POWER-ONE INC	23-Oct-2000	13-Mar-2005	5668	3679	Electronic Components, NEC
537	NI	NISOURCE INC	2-Nov-2000		5658	4932	Gas & Other Services Combined
538	EOG	EOG RESOURCES INC	2-Nov-2000		5658	1311	Crude Petroleum & Natural Gas
539	BVSN	BROADVISION INC	6-Nov-2000	3-Sep-2001	5654	7370	Services-Computer Programming, Data Processing, Etc.
540	HOT	STARWOOD HOTELS&RESORTS WRLD	17-Nov-2000	22-Sep-2016	5643	7011	Hotels & Motels
541	FRX	FOREST LABORATORIES -CL A	22-Nov-2000	30-Jun-2014	5638	2834	Pharmaceutical Preparations
542	CHIR	CHIRON CORP	24-Nov-2000	19-Apr-2006	5636	2834	Pharmaceutical Preparations
543	CPN	CALPINE CORP	1-Dec-2000	1-Dec-2005	5629	4991	Co-generation Services & Small Power Producers
544	RHI	ROBERT HALF INTL INC	5-Dec-2000		5625	7363	Services-Help Supply Services
545	INTU	INTUIT INC	11-Dec-2000		5619	7372	Services-Prepackaged Software
546	MET	METLIFE INC	11-Dec-2000		5619	6311	Life Insurance
547	SBL.2	SYMBOL TECHNOLOGIES	11-Dec-2000	9-Jan-2007	5619	3577	Computer Peripheral Equipment, NEC
548	AMBC	AMBAC FINANCIAL GROUP INC	11-Dec-2000	10-Jun-2008	5619	6351	Surety Insurance
549	AYE	ALLEGHENY ENERGY INC	11-Dec-2000	27-Feb-2011	5619	4911	Electric Services

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
550	SYK	STRYKER CORP	12-Dec-2000		5618	3842	Orthopedic, Prosthetic & Surgical Appliances & Supplies
551	VTSS	VITESSE SEMICONDUCTOR CORP	12-Dec-2000	20-Aug-2002	5618	3674	Semiconductors & Related Devices
552	KMI	KINDER MORGAN INC	12-Dec-2000	30-May-2007	5618	4923	Natural Gas Transmission & Distribution
553	QLGC	QLOGIC CORP	12-Dec-2000	17-Jan-2011	5618	3570	Computer & office Equipment
554	AMCC	APPLIED MICRO CIRCUITS CORP	2-Jan-2001	17-May-2006	5597	3674	Semiconductors & Related Devices
555	NE	NOBLE CORP PLC	16-Jan-2001	26-Mar-2009	5583	1381	Drilling Oil & Gas Wells
556	JBL	JABIL INC	30-Jan-2001	4-Nov-2014	5569	3672	Printed Circuit Boards
557	5952B	UNIVISION COMMUNICATIONS INC	7-Feb-2001	28-Mar-2007	5561	4833	Television Broadcasting Stations
558	FTR	FRONTIER COMMUNICATIONS CORP	27-Feb-2001	19-Mar-2017	5541	4813	Telephone Communications (No Radiotelephone)
559	CTAS	CINTAS CORP	1-Mar-2001		5539	2320	Men's & Boys' Furnishings, Work Clothing, & Allied Garments
560	FISV	FISERV INC	2-Apr-2001		5507	7374	Services-Computer Processing & Data Preparation
561	CE.2	CONCORD EFS INC	2-Apr-2001	26-Feb-2004	5507	6099	Functions Related To Depository Banking, NEC
562	GEN.3	GENON ENERGY INC	3-Apr-2001	16-Jul-2003	5506	4911	Electric Services

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
563	PBG	PEPSI BOTTLING GROUP INC	14-May-2001	28-Feb-2010	5465	2086	Bottled & Canned Soft Drinks & Carbonated Waters
564	MWW	MONSTER WORLDWIDE INC	4-Jun-2001	18-Dec-2011	5444	7370	Services-Computer Programming, Data Processing, Etc.
565	ZION	ZIONS BANCORPORATION	25-Jun-2001		5423	6020	
566	JHF	HANCOCK JOHN FINL SVCS INC	28-Jun-2001	28-Apr-2004	5420	6311	Life Insurance
567	COL	ROCKWELL COLLINS INC	2-Jul-2001		5416	3728	Aircraft Parts & Auxiliary Equipment, NEC
568	AWE.2	AT&T WIRELESS SERVICES INC	9-Jul-2001	26-Oct-2004	5409	4812	Radiotelephone Communications
569	PMCS	PMC-SIERRA INC	3-Aug-2001	1-Jul-2007	5384	3674	Semiconductors & Related Devices
570	FDO	FAMILY DOLLAR STORES	6-Aug-2001	8-Jul-2015	5381	5331	Retail-Variety Stores
571	ZBH	ZIMMER BIOMET HOLDINGS INC	7-Aug-2001		5380	3842	Orthopedic, Prosthetic & Surgical Appliances & Supplies
572	ABC	AMERISOURCEBERGEN CORP	30-Aug-2001		5357	5122	Wholesale-Drugs, Proprietaries & Druggists' Sundries
573	CIEN	CIENA CORP	30-Aug-2001	20-Dec-2009	5357	3661	Telephone & Telegraph Apparatus
574	XL	XL GROUP LTD	4-Sep-2001		5352	6351	Surety Insurance
575	IGT.1	INTL GAME TECHNOLOGY	4-Sep-2001	22-Jun-2014	5352	7990	Services-Miscellaneous Amusement & Recreation

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
576	IMNX	IMMUNEX CORP	21-Sep-2001	16-Jul-2002	5335	2836	Biological Products, (No Diagnostic Substances)
577	EOP	EQUITY OFFICE PROPERTIES TR	10-Oct-2001	11-Feb-2007	5316	6798	Real Estate Investment Trusts
578	TE	TECO ENERGY INC	10-Oct-2001	30-Jun-2016	5316	4931	Electric & Other Services Combined
579	HMA	HEALTH MANAGEMENT ASSOC	7-Nov-2001	1-Mar-2007	5288	8062	Services-General Medical & Surgical Hospitals, NEC
580	NVDA	NVIDIA CORP	30-Nov-2001		5265	3674	Semiconductors & Related Devices
581	EQR	EQUITY RESIDENTIAL	3-Dec-2001		5262	6798	Real Estate Investment Trusts
582	GENZ	GENZYME CORP	14-Dec-2001	3-Apr-2011	5251	2836	Biological Products, (No Diagnostic Substances)
583	JNY	JONES GROUP INC	17-Dec-2001	3-Mar-2009	5248	2330	Women's, Misses', and Juniors Outerwear
584	WAT	WATERS CORP	2-Jan-2002		5232	3826	Laboratory Analytical Instruments
585	PCL	PLUM CREEK TIMBER CO INC	17-Jan-2002	21-Feb-2016	5217	2400	Lumber & Wood Products (No Furniture)
586	CB	CHUBB LTD	30-Jan-2002	17-Jul-2008	5204	6331	Fire, Marine & Casualty Insurance
587	RATL	RATIONAL SOFTWARE CORP	1-Feb-2002	23-Feb-2003	5202	7370	Services-Computer Programming, Data Processing, Etc.

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
588	MI.1	MARSHALL & ILSLEY CORP	11-Feb-2002	5-Jul-2011	5192	6020	
589	FHN	FIRST HORIZON NATIONAL CORP	6-May-2002	23-Jun-2013	5108	6020	
590	TT.2	TRANE INC	13-May-2002	5-Jun-2008	5101	3585	Air-Cond & Warm Air Heatg Equip & Comm & Indl Refrig Equip
591	BJS.1	BJ SERVICES CO	15-May-2002	29-Apr-2010	5099	1389	Oil & Gas Field Services, NEC
592	APOL	APOLLO EDUCATION GROUP INC	15-May-2002	30-Jun-2013	5099	8200	Services-Educational Services
593	SPG	SIMON PROPERTY GROUP INC	26-Jun-2002		5057	6798	Real Estate Investment Trusts
594	NFB	NORTH FORK BANCORPORATION	17-Jul-2002	30-Nov-2006	5036	6020	
595	UPS	UNITED PARCEL SERVICE INC	22-Jul-2002		5031	4210	Trucking & Courier Services (No Air)
596	EA	ELECTRONIC ARTS INC	22-Jul-2002		5031	7372	Services-Prepackaged Software
597	EBAY	EBAY INC	22-Jul-2002		5031	7370	Services-Computer Programming, Data Processing, Etc.
598	GS	GOLDMAN SACHS GROUP INC	22-Jul-2002		5031	6211	Security Brokers, Dealers & Flotation Companies
599	PRU	PRUDENTIAL FINANCIAL INC	22-Jul-2002		5031	6311	Life Insurance
600	PFG	PRINCIPAL FINANCIAL GRP INC	22-Jul-2002		5031	6282	Investment Advice
601	0139A	SUNGARD DATA SYSTEMS INC	22-Jul-2002	11-Aug-2005	5031	7372	Services-Prepackaged Software
602	ANTM	ANTHEM INC	25-Jul-2002		5028	6324	Hospital & Medical Service Plans
603	MON	MONSANTO CO	14-Aug-2002	6-Jun-2018	5008	100	Agricultural Production-Crops

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
604	TRV	TRAVELERS COS INC	21-Aug-2002		5001	6331	Fire, Marine & Casualty Insurance
605	RAI	REYNOLDS AMERICAN INC	4-Sep-2002	25-Jul-2017	4987	2111	Cigarettes
606	CMCSA	COMCAST CORP	19-Nov-2002		4911	4841	Cable & Other Pay Television Services
607	DGX	QUEST DIAGNOSTICS INC	12-Dec-2002		4888	8071	Services-Medical Laboratories
608	AN	AUTONATION INC	24-Feb-2003	7-Aug-2017	4814	5500	Retail-Auto Dealers & Gasoline Stations
609	AIV	APARTMENT INVST & MGMT CO	14-Mar-2003		4796	6798	Real Estate Investment Trusts
610	MKC	MCCORMICK & CO INC	21-Mar-2003		4789	2090	Miscellaneous Food Preparations & Kindred Products
611	SYMC	SYMANTEC CORP	31-Mar-2003		4779	7372	Services-Prepackaged Software
612	FII	FEDERATED INVESTORS INC	16-Apr-2003	1-Jan-2013	4763	6282	Investment Advice
613	PLD	PROLOGIS INC	17-Jul-2003		4671	6798	Real Estate Investment Trusts
614	MHS	MEDCO HEALTH SOLUTIONS INC	20-Aug-2003	3-Apr-2012	4637	5912	Retail-Drug Stores and Proprietary Stores
615	ESRX	EXPRESS SCRIPTS HOLDING CO	26-Sep-2003		4600	5912	Retail-Drug Stores and Proprietary Stores
616	BIIB	BIOGEN INC	13-Nov-2003		4552	2836	Biological Products, (No Diagnostic Substances)
617	MTB	M & T BANK CORP	27-Feb-2004		4446	6020	

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
618	CMX.1	CAREMARK RX INC	25-Mar-2004	22-Mar-2007	4419	5912	Retail-Drug Stores and Proprietary Stores
619	ETFC	E TRADE FINANCIAL CORP	1-Apr-2004		4412	6211	Security Brokers, Dealers & Flotation Companies
620	ACS	AFFILIATED COMPUTER SERVICES	2-Apr-2004	7-Feb-2010	4411	7374	Services-Computer Processing & Data Preparation
621	MYL	MYLAN NV	23-Apr-2004		4390	2834	Pharmaceutical Preparations
622	VLO	VALERO ENERGY CORP	29-Apr-2004		4384	2911	Petroleum Refining
623	HSP	HOSPIRA INC	3-May-2004	2-Sep-2015	4380	2834	Pharmaceutical Preparations
624	GILD	GILEAD SCIENCES INC	1-Jul-2004		4321	2836	Biological Products, (No Diagnostic Substances)
625	STD2	SANTANDER HOLDINGS USA INC	1-Jul-2004	29-Jan-2009	4321	6020	
626	FSH.2	FISHER SCIENTIFIC INTL INC	3-Aug-2004	9-Nov-2006	4288	5040	Wholesale-Professional & Commercial Equipment & Supplies
627	TPR	TAPESTRY INC	1-Sep-2004		4259	3100	Leather & Leather Products
628	CIT	CIT GROUP INC	27-Oct-2004	26-Jul-2009	4203	6172	Finance Lessors
629	LH	LABORATORY CP OF AMER HLDGS	1-Nov-2004		4198	8071	Services-Medical Laboratories
630	LLL	L3 TECHNOLOGIES INC	1-Dec-2004		4168	3812	Search, Detection, Navigation, Guidance, Aeronautical Sys

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
631	BX1	FREESCALE SEMICONDUCTOR INC	3-Dec-2004	3-Dec-2006	4166	3674	Semiconductors & Related Devices
632	FOXA	TWENTY-FIRST CENTURY FOX INC	20-Dec-2004		4149	4888	
633	BBVA1	BBVA COMPASS BANCSHARES INC	20-Dec-2004	6-Sep-2007	4149	6020	
634	ASN	ARCHSTONE INC-REDH	20-Dec-2004	7-Oct-2007	4149	6798	Real Estate Investment Trusts
635	XTO	XTO ENERGY INC	29-Dec-2004	27-Jun-2010	4140	1311	Crude Petroleum & Natural Gas
636	NOV	NATIONAL OILWELL VARCO INC	14-Mar-2005		4065	3533	Oil & Gas Field Machinery & Equipment
637	SHLD	SEARS HOLDINGS CORP	28-Mar-2005	4-Sep-2012	4051	5311	Retail-Department Stores
638	STZ	CONSTELLATION BRANDS	5-Jul-2005		3952	2082	Malt Beverages
639	DHI	D R HORTON INC	5-Jul-2005		3952	1531	Operative Builders
640	WFT	WEATHERFORD INTL PLC	22-Jul-2005	25-Feb-2009	3935	1381	Drilling Oil & Gas Wells
641	TSN	TYSON FOODS INC -CL A	11-Aug-2005		3915	2011	Meat Packing Plants
642	VNO	VORNADO REALTY TRUST	12-Aug-2005		3914	6798	Real Estate Investment Trusts
643	MUR	MURPHY OIL CORP	15-Aug-2005	25-Jul-2017	3911	1311	Crude Petroleum & Natural Gas
644	PSA	PUBLIC STORAGE	19-Aug-2005		3907	6798	Real Estate Investment Trusts
645	CVH	COVENTRY HEALTH CARE INC	30-Aug-2005	8-May-2013	3896	6324	Hospital & Medical Service Plans
646	AMP	AMERIPRISE FINANCIAL INC	3-Oct-2005		3862	6211	Security Brokers, Dealers & Flotation Companies
647	LEN	LENNAR CORP	4-Oct-2005		3861	1531	Operative Builders

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
648	PDCO	PATTERSON COMPANIES INC	11-Oct-2005	18-Mar-2018	3854	5047	Wholesale-Medical, Dental & Hospital Equipment & Supplies
649	AMZN	AMAZON.COM INC	21-Nov-2005		3813	5961	Retail-Catalog & Mail-Order Houses
650	GNW	GENWORTH FINANCIAL INC	2-Dec-2005	17-Nov-2015	3802	6311	Life Insurance
651	SSP	EW SCRIPPS -CL A	20-Dec-2005	30-Jun-2008	3784	4833	Television Broadcasting Stations
652	VIAB	VIACOM INC	3-Jan-2006		3770	4833	Television Broadcasting Stations
653	WFM	WHOLE FOODS MARKET INC	3-Jan-2006	28-Aug-2017	3770	5411	Retail-Grocery Stores
654	EL	LAUDER (ESTEE) COS INC -CL A	5-Jan-2006		3768	2844	Perfumes, Cosmetics & Other Toilet Preparations
655	VRSN	VERISIGN INC	1-Feb-2006		3741	7370	Services-Computer Programming, Data Processing, Etc.
656	HAR	HARMAN INTERNATIONAL INDS	1-Feb-2006	15-Mar-2017	3741	3651	Household Audio & Video Equipment
657	BRL	BARR PHARMACEUTICALS INC	27-Feb-2006	22-Dec-2008	3715	2834	Pharmaceutical Preparations
658	CHK	CHESAPEAKE ENERGY CORP	3-Mar-2006	18-Mar-2018	3711	1311	Crude Petroleum & Natural Gas
659	KIM	KIMCO REALTY CORP	3-Apr-2006		3680	6798	Real Estate Investment Trusts
660	BXP	BOSTON PROPERTIES INC	3-Apr-2006		3680	6798	Real Estate Investment Trusts
661	GOOGL	ALPHABET INC	3-Apr-2006		3680	7370	Services-Computer Programming, Data Processing, Etc.
662	DF	DEAN FOODS CO	3-Apr-2006	23-May-2013	3680	2020	Dairy Products
663	SNDK	SANDISK CORP	20-Apr-2006	12-May-2016	3663	3572	Computer Storage Devices

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
664	LM	LEGG MASON INC	24-Apr-2006	1-Dec-2016	3659	6282	Investment Advice
665	EQ	EMBARQ CORP	18-May-2006	30-Jun-2009	3635	4813	Telephone Communications (No Radiotelephone)
666	JNPR	JUNIPER NETWORKS INC	2-Jun-2006		3620	3576	Computer Communications Equipment
667	CBH.1	COMMERCE BANCORP INC/NJ	6-Jun-2006	30-Mar-2008	3616	6020	
668	CNX	CNX RESOURCES CORPORATION	28-Jun-2006	3-Mar-2016	3594	1311	Crude Petroleum & Natural Gas
669	WIN	WINDSTREAM HOLDINGS INC	18-Jul-2006	6-Apr-2015	3574	4813	Telephone Communications (No Radiotelephone)
670	RLGY	REALOGY HOLDINGS CORP	1-Aug-2006	9-Apr-2007	3560	6531	Real Estate Agents & Managers (For Others)
671	WYND	WYNDHAM DESTINATIONS INC	1-Aug-2006	30-May-2018	3560	6531	Real Estate Agents & Managers (For Others)
672	CME	CME GROUP INC	11-Aug-2006		3550	6200	Security & Commodity Brokers, Dealers, Exchanges & Services
673	WU	WESTERN UNION CO	2-Oct-2006		3498	6099	Functions Related To Depository Banking, NEC
674	SII	SMITH INTERNATIONAL INC	2-Oct-2006	26-Aug-2010	3498	2890	Miscellaneous Chemical Products
675	CELG	CELGENE CORP	6-Nov-2006		3463	2834	Pharmaceutical Preparations
676	FIS	FIDELITY NATIONAL INFO SVCS	10-Nov-2006		3459	7374	Services-Computer Processing & Data Preparation
677	CBRE	CBRE GROUP INC	10-Nov-2006		3459	6500	Real Estate

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
678	CTSH	COGNIZANT TECH SOLUTIONS	17-Nov-2006		3452	7370	Services-Computer Programming, Data Processing, Etc.
679	BTU	PEABODY ENERGY CORP	20-Nov-2006	21-Sep-2014	3449	1220	Bituminous Coal & Lignite Mining
680	IAC	IAC/INTERACTIVECORP	1-Dec-2006	20-Aug-2008	3438	7370	Services-Computer Programming, Data Processing, Etc.
681	STR	QUESTAR CORP	1-Dec-2006	30-Jun-2010	3438	4923	Natural Gas Transmission & Distribution
682	DTV.2	DIRECTV	4-Dec-2006	28-Jul-2015	3435	4841	Cable & Other Pay Television Services
683	TEX	TEREX CORP	20-Dec-2006	5-Nov-2008	3419	3530	Construction, Mining & Materials Handling Machinery & Equip
684	SE.7	SPECTRA ENERGY CORP	3-Jan-2007	27-Feb-2017	3405	4923	Natural Gas Transmission & Distribution
685	ESV	ENSCO PLC	4-Jan-2007	22-Dec-2009	3404	1381	Drilling Oil & Gas Wells
686	AVB	AVALONBAY COMMUNITIES INC	10-Jan-2007		3398	6798	Real Estate Investment Trusts
687	RL	RALPH LAUREN CORP	2-Feb-2007		3375	2320	Men's & Boys' Furnishings, Work Clothing, & Allied Garments
688	VAR	VARIAN MEDICAL SYSTEMS INC	12-Feb-2007		3365	3845	Electromedical & Electrotherapeutic Apparatus

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
689	HCBK	HUDSON CITY BANCORP INC	15-Feb-2007	2-Nov-2015	3362	6035	Savings Institution, Federally Chartered
690	WEC3	INTEGRYS HOLDING INC	22-Feb-2007	30-Jun-2015	3355	4931	Electric & Other Services Combined
691	CHRW	C H ROBINSON WORLDWIDE INC	2-Mar-2007		3347	4213	Trucking (No Local)
692	HST	HOST HOTELS & RESORTS INC	20-Mar-2007		3329	6798	Real Estate Investment Trusts
693	DDR	DDR CORP	23-Mar-2007	29-Mar-2009	3326	6798	Real Estate Investment Trusts
694	ANF	ABERCROMBIE & FITCH -CL A	29-Mar-2007	22-Dec-2013	3320	5651	Retail-Family Clothing Stores
695	MDLZ	MONDELEZ INTERNATIONAL INC	2-Apr-2007		3316	2000	Food and Kindred Products
696	AIZ	ASSURANT INC	10-Apr-2007		3308	6351	Surety Insurance
697	SUNEQ	SUNEDISON INC	31-May-2007	18-Dec-2011	3257	3620	Electrical Industrial Apparatus
698	PCP	PRECISION CASTPARTS CORP	1-Jun-2007	31-Jan-2016	3256	3728	Aircraft Parts & Auxiliary Equipment, NEC
699	DFS	DISCOVER FINANCIAL SVCS	2-Jul-2007		3225	6141	Personal Credit Institutions
700	GGP	GGP INC	2-Jul-2007	12-Nov-2008	3225	6798	Real Estate Investment Trusts
701	COV	COVIDIEN PLC	2-Jul-2007	4-Jun-2009	3225	3845	Electromedical & Electrotherapeutic Apparatus
702	TEL	TE CONNECTIVITY LTD	2-Jul-2007	25-Jun-2009	3225	3678	Electronic Connectors
703	ACAS	AMERICAN CAPITAL LTD	9-Jul-2007	3-Mar-2009	3218	6797	
704	AKAM	AKAMAI TECHNOLOGIES INC	12-Jul-2007		3215	7370	Services-Computer Programming, Data Processing, Etc.
705	JEF	JEFFERIES FINANCIAL GRP INC	27-Aug-2007		3169	2011	Meat Packing Plants

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
706	MCHP	MICROCHIP TECHNOLOGY INC	7-Sep-2007		3158	3674	Semiconductors & Related Devices
707	ICE	INTERCONTINENTAL EXCHANGE	26-Sep-2007		3139	6200	Security & Commodity Brokers, Dealers, Exchanges & Services
708	ANDV	ANDEAVOR	27-Sep-2007		3138	2911	Petroleum Refining
709	TDC	TERADATA CORP	1-Oct-2007	18-Jun-2017	3134	7370	Services-Computer Programming, Data Processing, Etc.
710	EXPE	EXPEDIA GROUP INC	2-Oct-2007		3133	4700	Transportation Services
711	NBL	NOBLE ENERGY INC	8-Oct-2007		3127	1311	Crude Petroleum & Natural Gas
712	EXPD	EXPEDITORS INTL WASH INC	10-Oct-2007		3125	4731	Arrangement of Transportation of Freight & Cargo
713	NYX	NYSE EURONEXT	25-Oct-2007	12-Nov-2013	3110	6200	Security & Commodity Brokers, Dealers, Exchanges & Services
714	JEC	JACOBS ENGINEERING GROUP INC	26-Oct-2007		3109	1600	Heavy Construction Other Than Bldg Const - Contractors
715	TIE	TITANIUM METALS CORP	29-Oct-2007	23-Dec-2012	3106	3350	Rolling Drawing & Extruding of Nonferrous Metals
716	POM	PEPCO HOLDINGS INC	9-Nov-2007	29-Mar-2016	3095	4911	Electric Services
717	MTW	MANITOWOC CO	16-Nov-2007	31-Aug-2009	3088	3530	Construction, Mining & Materials Handling Machinery & Equip
718	AMT	AMERICAN TOWER CORP	19-Nov-2007		3085	6798	Real Estate Investment Trusts

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
719	GME	GAMESTOP CORP	14-Dec-2007	24-Apr-2016	3060	5734	Retail-Computer & Computer Software Stores
720	RRC	RANGE RESOURCES CORP	21-Dec-2007	17-Jun-2018	3053	1311	Crude Petroleum & Natural Gas
721	GHC	GRAHAM HOLDINGS CO	31-Dec-2007	21-Sep-2014	3043	8200	Services-Educational Services
722	TSS	TOTAL SYSTEM SERVICES INC	2-Jan-2008		3041	7389	Services-Business Services, NEC
723	CAM	CAMERON INTERNATIONAL CORP	29-Jan-2008	3-Apr-2016	3014	3533	Oil & Gas Field Machinery & Equipment
724	HCP	HCP INC	31-Mar-2008		2952	6798	Real Estate Investment Trusts
725	PM	PHILIP MORRIS INTERNATIONAL	31-Mar-2008		2952	2111	Cigarettes
726	ISRG	INTUITIVE SURGICAL INC	2-Jun-2008		2889	3845	Electromedical & Electrotherapeutic Apparatus
727	SWN	SOUTHWESTERN ENERGY CO	6-Jun-2008	3-Apr-2017	2885	4923	Natural Gas Transmission & Distribution
728	LO	LORILLARD INC	11-Jun-2008	11-Jun-2015	2880	2111	Cigarettes
729	COG	CABOT OIL & GAS CORP	23-Jun-2008		2868	1311	Crude Petroleum & Natural Gas
730	MEE	MASSEY ENERGY CO	23-Jun-2008	1-Jun-2011	2868	1220	Bituminous Coal & Lignite Mining
731	AKS	AK STEEL HOLDING CORP	1-Jul-2008	18-Dec-2011	2860	3312	Steel Works, Blast Furnaces & Rolling Mills (Coke Ovens)
732	SNI	SCRIPPS NETWORKS INTERACTIVE	1-Jul-2008	6-Mar-2018	2860	4833	Television Broadcasting Stations

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
733	MA	MASTERCARD INC	18-Jul-2008		2843	6099	Functions Related To Depository Banking, NEC
734	DVA	DAVITA INC	31-Jul-2008		2830	8090	Services-Misc Health & Allied Services, NEC
735	IVZ	INVESCO LTD	21-Aug-2008		2809	6282	Investment Advice
736	CF	CF INDUSTRIES HOLDINGS INC	27-Aug-2008		2803	2870	Agricultural Chemicals
737	FAST	FASTENAL CO	15-Sep-2008		2784	5000	Wholesale-Durable Goods
738	CRM	SALESFORCE.COM INC	15-Sep-2008		2784	7372	Services-Prepackaged Software
739	HRS	HARRIS CORP	22-Sep-2008		2777	3812	Search, Detection, Navigation, Guidance, Aeronautical Sys
740	PXD	PIONEER NATURAL RESOURCES CO	24-Sep-2008		2775	1311	Crude Petroleum & Natural Gas
741	APH	AMPHENOL CORP	30-Sep-2008		2769	3678	Electronic Connectors
742	FLS	FLOWSERVE CORP	2-Oct-2008		2767	3561	Pumps & Pumping Equipment
743	DPS	DR PEPPER SNAPPLE GROUP INC	7-Oct-2008		2762	2086	Bottled & Canned Soft Drinks & Carbonated Waters
744	NDAQ	NASDAQ INC	22-Oct-2008		2747	6200	Security & Commodity Brokers, Dealers, Exchanges & Services
745	WEC	WEC ENERGY GROUP INC	31-Oct-2008		2738	4931	Electric & Other Services Combined
746	SJM	SMUCKER (JM) CO	6-Nov-2008		2732	2033	Canned, Fruits, Veg, Preserves, Jams & Jellies

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
747	PBCT	PEOPLE'S UNITED FINL INC	13-Nov-2008		2725	6020	
748	XRAY	DENTSPLY SIRONA INC	14-Nov-2008		2724	3843	Dental Equipment & Supplies
749	WYNN	WYNN RESORTS LTD	14-Nov-2008		2724	7990	Services-Miscellaneous Amusement & Recreation
750	CEPH	CEPHALON INC	17-Nov-2008	16-Oct-2011	2721	2834	Pharmaceutical Preparations
751	SRCL	STERICYCLE INC	19-Nov-2008		2719	4955	Hazardous Waste Management
752	LIFE.3	LIFE TECHNOLOGIES CORP	24-Nov-2008	23-Jan-2014	2714	2836	Biological Products, (No Diagnostic Substances)
753	DNB	DUN & BRADSTREET CORP	2-Dec-2008	4-Apr-2017	2706	7323	
754	RSG	REPUBLIC SERVICES INC	5-Dec-2008		2703	4953	Refuse Systems
755	EQT	EQT CORP	19-Dec-2008		2689	4923	Natural Gas Transmission & Distribution
756	MFE	MCAFEE INC	23-Dec-2008	28-Feb-2011	2685	7372	Services-Prepackaged Software
757	SCG	SCANA CORP	2-Jan-2009		2675	4931	Electric & Other Services Combined
758	FLIR	FLIR SYSTEMS INC	2-Jan-2009		2675	3812	Search, Detection, Navigation, Guidance, Aeronautical Sys
759	OI	OWENS-ILLINOIS INC	2-Jan-2009	1-Dec-2016	2675	3221	Glass Containers
760	IRM	IRON MOUNTAIN INC	6-Jan-2009		2671	6798	Real Estate Investment Trusts
761	WELL	WELLTOWER INC	30-Jan-2009		2647	6798	Real Estate Investment Trusts
762	DO	DIAMOND OFFSHRE DRILLING INC	26-Feb-2009	2-Oct-2016	2620	1381	Drilling Oil & Gas Wells

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
763	HRL	HORMEL FOODS CORP	4-Mar-2009		2614	2011	Meat Packing Plants
764	VTR	VENTAS INC	4-Mar-2009		2614	6798	Real Estate Investment Trusts
765	ES	EVERSOURCE ENERGY	17-Mar-2009		2601	4931	Electric & Other Services Combined
766	ORLY	O'REILLY AUTOMOTIVE INC	27-Mar-2009		2591	5531	Retail-Auto & Home Supply Stores
767	TWC	TIME WARNER CABLE INC	30-Mar-2009	17-May-2016	2588	4841	Cable & Other Pay Television Services
768	DNR	DENBURY RESOURCES INC	2-Apr-2009	22-Mar-2015	2585	1311	Crude Petroleum & Natural Gas
769	FTI.1	FMC TECHNOLOGIES INC	5-Jun-2009	16-Jan-2017	2521	3533	Oil & Gas Field Machinery & Equipment
770	ATGE	ADTALEM GLOBAL EDUCATION INC	9-Jun-2009	30-Sep-2012	2517	8200	Services-Educational Services
771	PCS	METROPCS COMMUNICATIONS INC	30-Jun-2009	30-Apr-2013	2496	4812	Radiotelephone Communications
772	WDC	WESTERN DIGITAL CORP	1-Jul-2009		2495	3572	Computer Storage Devices
773	PWR	QUANTA SERVICES INC	1-Jul-2009		2495	1731	Electrical Work
774	RHT	RED HAT INC	27-Jul-2009		2469	7372	Services-Prepackaged Software
775	FMC	FMC CORP	19-Aug-2009		2446	2870	Agricultural Chemicals
776	CFN	CAREFUSION CORP	1-Sep-2009	17-Mar-2015	2433	3841	Surgical & Medical Instruments & Apparatus

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
777	ARG	AIRGAS INC	9-Sep-2009	22-May-2016	2425	5084	Wholesale-Industrial Machinery & Equipment
778	FSLR	FIRST SOLAR INC	16-Oct-2009	19-Mar-2017	2388	3674	Semiconductors & Related Devices
779	BKNG	BOOKING HOLDINGS INC	6-Nov-2009		2367	7370	Services-Computer Programming, Data Processing, Etc.
780	ROST	ROSS STORES INC	21-Dec-2009		2322	5651	Retail-Family Clothing Stores
781	V	VISA INC	21-Dec-2009		2322	6099	Functions Related To Depository Banking, NEC
782	LDOS	LEIDOS HOLDINGS INC	21-Dec-2009	22-Sep-2013	2322	7373	Services-Computer Integrated Systems Design
783	CLF	CLEVELAND-CLIFFS INC	21-Dec-2009	1-Apr-2014	2322	1000	Metal Mining
784	MJN	MEAD JOHNSON NUTRITION CO	21-Dec-2009	18-Jun-2017	2322	2020	Dairy Products
785	ROP	ROPER TECHNOLOGIES INC	23-Dec-2009		2320	3823	Industrial Instruments For Measurement, Display, and Control
786	NRG	NRG ENERGY INC	29-Jan-2010		2283	4911	Electric Services
787	URBN	URBAN OUTFITTERS INC	8-Feb-2010	19-Mar-2017	2273	5651	Retail-Family Clothing Stores
788	BRK.B	BERKSHIRE HATHAWAY	16-Feb-2010		2265	9997	
789	HP	HELMERICH & PAYNE	1-Mar-2010		2252	1381	Drilling Oil & Gas Wells
790	DISCA	DISCOVERY INC	1-Mar-2010		2252	4841	Cable & Other Pay Television Services

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
791	OKE	ONEOK INC	15-Mar-2010		2238	4923	Natural Gas Transmission & Distribution
792	CERN	CERNER CORP	30-Apr-2010		2192	7373	Services-Computer Integrated Systems Design
793	KMX	CARMAX INC	28-Jun-2010		2133	5500	Retail-Auto Dealers & Gasoline Stations
794	QEP	QEP RESOURCES INC	1-Jul-2010	30-Jun-2015	2130	1311	Crude Petroleum & Natural Gas
795	CB	CHUBB LTD	15-Jul-2010		2116	6331	Fire, Marine & Casualty Insurance
796	TYC	TYCO INTERNATIONAL PLC	27-Aug-2010	5-Sep-2016	2073	3669	Communications Equipment, NEC
797	IR	INGERSOLL-RAND PLC	17-Nov-2010		1991	3585	Air-Cond & Warm Air Heatg Equip & Comm & Indl Refrig Equip
798	NFX	NEWFIELD EXPLORATION CO	20-Dec-2010		1958	1311	Crude Petroleum & Natural Gas
799	FFIV	F5 NETWORKS INC	20-Dec-2010		1958	7373	Services-Computer Integrated Systems Design
800	NFLX	NETFLIX INC	20-Dec-2010		1958	7841	Services-Video Tape Rental
801	ATUS	ALTICE USA INC	20-Dec-2010	23-Jun-2016	1958	4841	Cable & Other Pay Television Services
802	MMI.3	MOTOROLA MOBILITY HLDGS INC	4-Jan-2011	24-May-2012	1943	3663	Radio & TV Broadcasting & Communications Equipment

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
803	NE	NOBLE CORP PLC	18-Jan-2011	19-Jul-2015	1929	1381	Drilling Oil & Gas Wells
804	JOY	JOY GLOBAL INC	28-Feb-2011	7-Oct-2015	1888	3532	Mining Machinery & Equip (No Oil & Gas Field Mach & Equip)
805	COV	COVIDIEN PLC	1-Mar-2011	26-Jan-2015	1887	3845	Electromedical & Electrotherapeutic Apparatus
806	EW	EDWARDS LIFESCIENCES CORP	1-Apr-2011		1856	3842	Orthopedic, Prosthetic & Surgical Appliances & Supplies
807	BLK	BLACKROCK INC	4-Apr-2011		1853	6282	Investment Advice
808	CMG	CHIPOTLE MEXICAN GRILL INC	28-Apr-2011		1829	5812	Retail-Eating Places
809	ANRZQ	ALPHA NATURAL RESOURCES INC	2-Jun-2011	1-Oct-2012	1794	1220	Bituminous Coal & Lignite Mining
810	MPC	MARATHON PETROLEUM CORP	1-Jul-2011		1765	2911	Petroleum Refining
811	ACN	ACCENTURE PLC	6-Jul-2011		1760	8742	Services-Management Consulting Services
812	MOS	MOSAIC CO	26-Sep-2011		1678	2870	Agricultural Chemicals
813	TEL	TE CONNECTIVITY LTD	17-Oct-2011		1657	3678	Electronic Connectors
814	XYL	XYLEM INC	1-Nov-2011		1642	3561	Pumps & Pumping Equipment
815	CBE	COOPER INDUSTRIES PLC	23-Nov-2011	2-Dec-2012	1620	3640	Electric Lighting & Wiring Equipment
816	SO7	SOUTHERN CO GAS	13-Dec-2011	30-Jun-2016	1600	4924	Natural Gas Distribution
817	PRGO	PERRIGO CO PLC	19-Dec-2011		1594	2834	Pharmaceutical Preparations

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
818	BWA	BORGWARNER INC	19-Dec-2011		1594	3714	Motor Vehicle Parts & Accessories
819	DLTR	DOLLAR TREE INC	19-Dec-2011		1594	5331	Retail-Variety Stores
820	TRIP	TRIPADVISOR INC	21-Dec-2011		1592	7370	Services-Computer Programming, Data Processing, Etc.
821	WPX	WPX ENERGY INC	3-Jan-2012	23-Mar-2014	1579	1311	Crude Petroleum & Natural Gas
822	CCI	CROWN CASTLE INTL CORP	14-Mar-2012		1508	6798	Real Estate Investment Trusts
823	FOSL	FOSSIL GROUP INC	4-Apr-2012	4-Jan-2016	1487	3873	Watches, Clocks, Clockwork Operated Devices/Parts
824	PSX	PHILLIPS 66	1-May-2012		1460	2911	Petroleum Refining
825	KMI	KINDER MORGAN INC	25-May-2012		1436	4923	Natural Gas Transmission & Distribution
826	ALXN	ALEXION PHARMACEUTICALS INC	25-May-2012		1436	2836	Biological Products, (No Diagnostic Substances)
827	LRCX	LAM RESEARCH CORP	5-Jun-2012		1425	3559	Special Industry Machinery, NEC
828	MNST	MONSTER BEVERAGE CORP	29-Jun-2012		1401	2086	Bottled & Canned Soft Drinks & Carbonated Waters
829	STX	SEAGATE TECHNOLOGY PLC	2-Jul-2012		1398	3572	Computer Storage Devices
830	ESV	ENSCO PLC	31-Jul-2012	29-Mar-2016	1369	1381	Drilling Oil & Gas Wells
831	LYB	LYONDELLBASELL INDUSTRIES NV	5-Sep-2012		1333	2820	Plastic Material, Synth Resin/Rubber, Cellulos (No Glass)

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
832	PNR	PENTAIR PLC	1-Oct-2012		1307	3561	Pumps & Pumping Equipment
833	ADT.4	ADT CORP	1-Oct-2012	2-May-2016	1307	7380	Services-Miscellaneous Business Services
834	KRFT	KRAFT FOODS GROUP INC	2-Oct-2012	2-Jul-2015	1306	2000	Food and Kindred Products
835	PETM	PETSMART INC	5-Oct-2012	11-Mar-2015	1303	5990	Retail-Retail Stores, NEC
836	DG	DOLLAR GENERAL CORP	3-Dec-2012		1244	5331	Retail-Variety Stores
837	GRMN	GARMIN LTD	12-Dec-2012		1235	3812	Search, Detection, Navigation, Guidance, Aeronautical Sys
838	APTV	APTIV PLC	24-Dec-2012		1223	3714	Motor Vehicle Parts & Accessories
839	ABBV	ABBVIE INC	2-Jan-2013		1214	2836	Biological Products, (No Diagnostic Substances)
840	PVH	PVH CORP	14-Feb-2013		1171	2300	Apparel & Other Finished Prods of Fabrics & Similar Matl
841	REGN	REGENERON PHARMACEUTICALS	1-May-2013		1095	2834	Pharmaceutical Preparations
842	MAC	MACERICH CO	9-May-2013		1087	6798	Real Estate Investment Trusts
843	KSU	KANSAS CITY SOUTHERN	24-May-2013		1072	4011	Railroads, Line-Haul Operating
844	GM	GENERAL MOTORS CO	7-Jun-2013		1058	3711	Motor Vehicles & Passenger Car Bodies
845	ZTS	ZOETIS INC	24-Jun-2013		1041	2834	Pharmaceutical Preparations

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
846	NWSA	NEWS CORP	1-Jul-2013		1034	2711	Newspapers: Publishing or Publishing & Printing
847	NLSN	NIELSEN HOLDINGS PLC	9-Jul-2013		1026	8700	Services-Engineering, Accounting, Research, Management
848	DAL	DELTA AIR LINES INC	11-Sep-2013		962	4512	Air Transportation, Scheduled
849	AME	AMETEK INC	23-Sep-2013		950	3823	Industrial Instruments For Measurement, Display, and Control
850	VRTX	VERTEX PHARMACEUTICALS INC	23-Sep-2013		950	2834	Pharmaceutical Preparations
851	RIG	TRANSOCEAN LTD	29-Oct-2013	25-Jul-2017	914	1381	Drilling Oil & Gas Wells
852	KORS	MICHAEL KORS HOLDINGS LTD	13-Nov-2013		899	2300	Apparel & Other Finished Prods of Fabrics & Similar Matl
853	ALLE	ALLEGION PLC	2-Dec-2013		880	3420	Cutlery, Handtools & General Hardware
854	GGP	GGP INC	10-Dec-2013		872	6798	Real Estate Investment Trusts
855	MHK	MOHAWK INDUSTRIES INC	23-Dec-2013		859	2273	Carpets & Rugs
856	ADS	ALLIANCE DATA SYSTEMS CORP	23-Dec-2013		859	7374	Services-Computer Processing & Data Preparation
857	FB	FACEBOOK INC	23-Dec-2013		859	7370	Services-Computer Programming, Data Processing, Etc.

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
858	TSCO	TRACTOR SUPPLY CO	24-Jan-2014		827	5200	Retail-Building Materials, Hardware, Garden Supply
859	GMCR	KEURIG GREEN MOUNTAIN INC	24-Mar-2014	6-Mar-2016	768	2090	Miscellaneous Food Preparations & Kindred Products
860	ESS	ESSEX PROPERTY TRUST	2-Apr-2014		759	6798	Real Estate Investment Trusts
861	GOOGL	ALPHABET INC	3-Apr-2014		758	7370	Services-Computer Programming, Data Processing, Etc.
862	UAA	UNDER ARMOUR INC	1-May-2014		730	2300	Apparel & Other Finished Prods of Fabrics & Similar Matl
863	AVGO	BROADCOM INC	8-May-2014		723	3674	Semiconductors & Related Devices
864	XEC	CIMAREX ENERGY CO	23-Jun-2014		677	1311	Crude Petroleum & Natural Gas
865	AMG	AFFILIATED MANAGERS GRP INC	1-Jul-2014		669	6282	Investment Advice
866	MLM	MARTIN MARIETTA MATERIALS	2-Jul-2014		668	1400	Mining & Quarrying of Nonmetallic Minerals (No Fuels)
867	DISCA	DISCOVERY INC	7-Aug-2014		632	4841	Cable & Other Pay Television Services
868	MNK	MALLINCKRODT PLC	19-Aug-2014	25-Jul-2017	620	2834	Pharmaceutical Preparations
869	UHS	UNIVERSAL HEALTH SVCS INC	22-Sep-2014		586	8062	Services-General Medical & Surgical Hospitals, NEC

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
870	URI	UNITED RENTALS INC	22-Sep-2014		586	7350	Services-Miscellaneous Equipment Rental & Leasing
871	LVLT	LEVEL 3 COMMUNICATIONS INC	5-Nov-2014	12-Oct-2017	542	4813	Telephone Communications (No Radiotelephone)
872	RCL	ROYAL CARIBBEAN CRUISES LTD	5-Dec-2014		512	4400	Water Transportation
873	HCA	HCA HEALTHCARE INC	27-Jan-2015		459	8062	Services-General Medical & Surgical Hospitals, NEC
874	ENDP	ENDO INTERNATIONAL PLC	27-Jan-2015	1-Mar-2017	459	2834	Pharmaceutical Preparations
875	SWKS	SKYWORKS SOLUTIONS INC	12-Mar-2015		415	3674	Semiconductors & Related Devices
876	HSIC	SCHEIN (HENRY) INC	18-Mar-2015		409	5047	Wholesale-Medical, Dental & Hospital Equipment & Supplies
877	AAL	AMERICAN AIRLINES GROUP INC	23-Mar-2015		404	4512	Air Transportation, Scheduled
878	SLG	SL GREEN REALTY CORP	23-Mar-2015		404	6798	Real Estate Investment Trusts
879	EQIX	EQUINIX INC	23-Mar-2015		404	6798	Real Estate Investment Trusts
880	HBI	HANESBRANDS INC	23-Mar-2015		404	2300	Apparel & Other Finished Prods of Fabrics & Similar Matl
881	O	REALTY INCOME CORP	7-Apr-2015		389	6798	Real Estate Investment Trusts
882	QRVO	QORVO INC	12-Jun-2015		323	3674	Semiconductors & Related Devices
883	JBHT	HUNT (JB) TRANSPRT SVCS INC	1-Jul-2015		304	4213	Trucking (No Local)

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
884	BXLT	BAXALTA INC	1-Jul-2015	2-Jun-2016	304	2836	Biological Products, (No Diagnostic Substances)
885	WRK	WESTROCK CO	2-Jul-2015		303	2650	Paperboard Containers & Boxes
886	CPGX	COLUMBIA PIPELINE GROUP INC	2-Jul-2015	5-Jul-2016	303	4922	Natural Gas Transmission
887	KHC	KRAFT HEINZ CO	3-Jul-2015		302	2030	Canned, Frozen & Preserved Fruit, Veg & Food Specialties
888	AAP	ADVANCE AUTO PARTS INC	9-Jul-2015		296	5531	Retail-Auto & Home Supply Stores
889	PYPL	PAYPAL HOLDINGS INC	20-Jul-2015		285	7374	Services-Computer Processing & Data Preparation
890	SIG	SIGNET JEWELERS LTD	29-Jul-2015	18-Mar-2018	276	5944	Retail-Jewelry Stores
891	ATVI	ACTIVISION BLIZZARD INC	31-Aug-2015		243	7372	Services-Prepackaged Software
892	UAL	UNITED CONTINENTAL HLDGS INC	3-Sep-2015		240	4512	Air Transportation, Scheduled
893	FOXA	TWENTY-FIRST CENTURY FOX INC	21-Sep-2015		222	4888	
894	NWSA	NEWS CORP	21-Sep-2015		222	2711	Newspapers: Publishing or Publishing & Printing
895	CMCSA	COMCAST CORP	21-Sep-2015	13-Dec-2015	222	4841	Cable & Other Pay Television Services
896	VRSK	VERISK ANALYTICS INC	8-Oct-2015		205	6411	Insurance Agents, Brokers & Service

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
897	HPE	HEWLETT PACKARD ENTERPRISE	2-Nov-2015		180	3571	Electronic Computers
898	FCPT	FOUR CORNERS PROPERTY TR INC	10-Nov-2015	10-Nov-2015	172	6798	Real Estate Investment Trusts
899	SYF	SYNCHRONY FINANCIAL	18-Nov-2015		164	6141	Personal Credit Institutions
900	ILMN	ILLUMINA INC	19-Nov-2015		163	3826	Laboratory Analytical Instruments
901	CSRA	CSRA INC	30-Nov-2015	3-Apr-2018	152	7370	Services-Computer Programming, Data Processing, Etc.
902	CHD	CHURCH & DWIGHT INC	29-Dec-2015		123	2840	Soap, Detergents, Cleaning Preparations, Perfumes, Cosmetics
903	WLTW	WILLIS TOWERS WATSON PLC	5-Jan-2016		116	6411	Insurance Agents, Brokers & Service
904	EXR	EXTRA SPACE STORAGE INC	19-Jan-2016		102	6798	Real Estate Investment Trusts
905	FRT	FEDERAL REALTY INVESTMENT TR	1-Feb-2016		89	6798	Real Estate Investment Trusts
906	CFG	CITIZENS FINANCIAL GROUP INC	1-Feb-2016		89	6020	
907	CXO	CONCHO RESOURCES INC	22-Feb-2016		68	1311	Crude Petroleum & Natural Gas
908	AWK	AMERICAN WATER WORKS CO INC	4-Mar-2016		57	4941	Water Supply
909	UDR	UDR INC	7-Mar-2016		54	6798	Real Estate Investment Trusts

Continued on next page

Idx	Ticker	Company Name	Date First Added	Date Last Appear	Total Days	Company SIC Code	Industry
910	HOLX	HOLOGIC INC	30-Mar-2016		31	3844	X-Ray Apparatus & Tubes & Related Irradiation Apparatus
911	CNC	CENTENE CORP	30-Mar-2016		31	6324	Hospital & Medical Service Plans
912	FL	FOOT LOCKER INC	4-Apr-2016		26	5661	Retail-Shoe Stores
913	UAA	UNDER ARMOUR INC	8-Apr-2016		22	2300	Apparel & Other Finished Prods of Fabrics & Similar Matl
914	ULTA	ULTA BEAUTY INC	18-Apr-2016		12	5990	Retail-Retail Stores, NEC
915	GPN	GLOBAL PAYMENTS INC	25-Apr-2016		5	7374	Services-Computer Processing & Data Preparation